NUCLEIC ACID DEGRADATION BY THE CRISPR TYPE III-B COMPLEX

by
Michael Anthony Estrella

A dissertation submitted to Johns Hopkins University in conformity with the

requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
April 2016

Abstract

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) system is an adaptive immune system utilized by prokaryotes to preserve their genome against invasive nucleic acid. CRISPR-derived RNA (crRNA) and CRISPR associated (Cas) proteins form an effector complex that degrades foreign genetic elements and is the basis of this prokaryotic defense system. The CRISPR Type III-B effector complex, also referred to as the Cmr complex, is composed of 6 protein subunits (Cmr1-6) and a crRNA. It has been shown to cleave ssRNA *in vitro* and degrade transcriptionally active DNA *in vivo*. To understand the mechanism of this activity, we undertook a biochemical analysis of the Cmr complex to determine how it cleaves nucleic acid.

We have been able to recombinantly express and purify the individual Cmr subunits of *Thermotoga maritima MSB8*, to form the Cmr complex with a short crRNA *in vitro*. Using pull-down assays we demonstrate that Cmr proteins from *Thermotoga maritima* assemble into a Cmr complex with the same subunit interactions as those observed for Cmr complexes from other species. Furthermore, we have reconstituted CRISPR-mediated degradation of RNA *in vitro*. Analysis of this activity demonstrates the preference for cleavage of ssRNA at 6 nucleotide intervals and the dependence on the 2' hydroxyl adjacent to the scissile phosphate for cleavage.

We also report ssDNA cleavage activity by the Cmr complex, which was previously unknown. Cleavage is observed only at thymidine nucleotides and is dependent upon the presence of target RNA complementary to the crRNA of the complex. We also demonstrate that as complementary target RNA is degraded, DNA activity is turned off. We mapped the active site for ssDNA cleavage to the HD domain of the Cmr2 subunit.

Our *in vitro* results are consistent with transcription-coupled DNA targeting by the Cmr complex. This study further resolves the ambiguity in activities reported for Type III systems and supports a unified mechanism for all Type III systems in which DNA nuclease activity is coupled to transcription.

Thesis Advisor: Dr. Scott Bailey

Thesis Readers: Dr. Sean Prigge, Dr. Paul Miller, Dr. Herschel Wade, and Dr. Andrew Pekosz

Alternate Readers: Dr. Roger McMacken and Dr. Cynthia Wolberger

Preface

There have been those that have guided me and taught me science and those that have supported and enabled me to pursue my love of it. I am so fortunate for all these people I have met on my journey to and through graduate school, and I hope they understand how much they mean to me.

First and foremost, I would like to thank my advisor, Scott Bailey, for accepting me into his lab and guiding me through a challenging project. I will always cherish our stimulating scientific conversations lasting hours and appreciate your patience with me. I am also indebted to my thesis committee members: Roger McMacken, Paul Miller, Sean Prigge and Herschel Wade. Thank you for your support throughout my graduate career, and for the myriad conversations about science and life after graduate school.

I would also like to thank Brian Learn and past lab members Gabriel Brandt and Sabin Mulepati for their advice on experiments and presentations. I am thankful to Jürgen Bosch, Michael Matunis and Pierre Coulombe for allowing me to use their equipment in times of need. I am appreciative to all the members of the Bailey and Bosch labs over the years that have enriched my social and professional life and have contributed to my work. I am also thankful to the members of my BMB class (Cathy Chen, Daisy Colón-López, Shaina Palmere, Eric Wier, Lauren McGinnis, Dolly Singh and Casey Daniels) for helping me get through our first year classes and for numerous discussions about

experiments and careers. In addition, I would like to thank Cathy for being an awesome lab partner, Daisy and Shaina for all the laughs and Eric for being a great roommate.

There are so many wonderful people that helped me get to Johns Hopkins. Most importantly is my undergraduate advisor, Filomena Califano. She taught me to be passionate about science and proffered so many opportunities for my scientific advancement. Then there is José Rodriguez, who gave me my first opportunity to work in a lab professionally and is the most productive person I know. I would like to thank Terrence Buck who made it all happen and Laura Barrio who is great friend and inspiration for me.

Last, I would like to profusely thank my family for supporting me. Thank you to my mother, Candida Vasquez, for telling me I can do anything I want and saying it with enough conviction to keep me motivated. Thanks to my older siblings Grace and John that have served as great role models and have given me awesome nieces and nephews. I would also like to thank my fiancé's mother, Maria Meregildo, and brother's Ismael and Fernando Caraballo for their love and support. Finally, I can't imagine graduate school let alone life without my fiancé, Lilibeth Caraballo. She has been my rock through all the ups and downs and the one person I could count on in times of need. Her unwavering love has helped me through it all and I am forever grateful for all her help. Thank you.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

Genome diversification and preservation in prokaryotes

Prokaryotes are exceptionally diverse, and attempts to quantify their diversity is understood to be beyond practical calculation (Doolittle & Zhaxybayeva 2009; Curtis et al. 2002; Whitman et al. 1998). The diversity of bacteria and archaea has allowed them to colonize every habitat and represent the largest component of biomass on earth (Suttle 2005). This surfeit of genetic variability is partially attributed to their ability to exchange genetic information through a process known as horizontal gene transfer (HGT). HGT, specifically involves swapping of genetic material by organisms that are not in a parent-progeny relationship (Soucy et al. 2015). In prokaryotes, the three best-characterized mechanisms of HGT are conjugation, transformation, and transduction (Thomas & Nielsen 2005). Conjugation requires contact between a donor and recipient cell with a pilus, through which genetic material is exchanged. Transformation is the absorption of extracellular genetic material from the environment. Lastly, transduction is the delivery of genetic material into cells via infection by bacteriophages. For example, the hyperthermophilic bacteria *Thermotoga maritima* is known to contain >20% open reading frames that are archaeal in origin (Nelson et al. 1999). To account for this observation, it is presumed that *Thermotoga maritima* acquired these genes through HGT from archaeal taxa subsequent to the Archaea–Bacteria split.

For transferred genes to successfully integrate, a requisite is that they do not harm the recipient. The acquired genetic information may be beneficial or neutral. If neutral, it may remain so and be lost over time or it may prove to be beneficial later on due to selection. Equally, this genetic information could be detrimental to the individual host and/or the entire colony. A major reservoir of harmful genetic information are viruses, which outnumber prokaryotes and kill up to 50% of the bacteria produced everyday (Wommack & Colwell 2000; Breitbart & Rohwer 2005). Due to the incessant threat viruses prose, prokaryotes have developed various mechanisms to impede the integration and propagation of harmful genetic information. These mechanisms neutralize the virus's effort to assume their parasitic relationship with prokaryotes at every step, including: attachment, nucleic acid injection, replication, transcription, viral particle assembly and lysis (Figure 1.1A) (Seed 2015). Specifically, the mechanisms are absorption inhibition, restriction-modification, abortive infection and CRISPR (Figure 1.1B) (Forde & Fitzgerald 1999; Bickle & Krüger 1993; Chopin et al. 2005; Rath et al. 2015).

Of the defense systems listed above, two specifically target invasive nucleic acid: restriction-modification (R-M) and CRISPR. Classically, R-M systems are comprised of a restriction endonuclease (RE) and a cognate methyltransferase. The methyltransferase will methylate self-DNA at specific sequences, whereas foreign DNA is unmodified. The basis of this system is methylation status at specific DNA sequences due to the REs activity of degrading unmethylated DNA at specific sites (Tock & Dryden 2005). The R-M system is supplemented by CRISPR to degrade foreign nucleic and protect prokaryotes

from harmful genetic elements. CRISPR, which stands for Clustered Regularly Interspaced Short Palindromic Repeats, is a recently discovered prokaryotic immune system that uses ribonucleoprotein complexes to bind specifically to foreign nucleic acid sequences and degrade them.

The CRISPR immune system

CRISPR is found in 87% of archaeal and 50% of bacterial species that have been sequenced (Makarova et al. 2015). The hallmarks of a CRISPR system in prokaryotic genomes are arrays of short repeating DNA sequences separated by similarly sized variable spacer sequences (Mojica et al. 2000). These repeating sequences were first identified in the genome of *Escherichia coli* in 1987, yet their function was unknown (Ishino et al. 1987). As more prokaryotic genomes were sequenced, the prevalence of these repeats became evident. In addition, it was noted that the variable spacer sequences were homologous to DNA from plasmids and viruses (Mojica et al. 2005; Pourcel et al. 2005; Bolotin et al. 2005). This observation implied the spacers were of extra-chromosomal origin and could be a genetic record of past infections. Furthermore, these sequences could serve as an immune system by preventing foreign DNA expression (Bolotin et al. 2005). In 2007, the first direct proof of this postulated immune activity was confirmed in *Streptococcus thermophilus* (Barrangou et al. 2007). It was demonstrated that upon bacteriophage infection *S. thermophilus* acquired new spacer sequences that were complementary to the bacteriophage genome.

With the function of CRISPR established, studies were then conducted to understand the molecular mechanism by which the system operates. Using *in silico* analysis it was observed that CRISPR loci were flanked by genes unique only to prokaryotes containing CRISPR (Jansen et al. 2002). These first CRISPR-associated (Cas) genes, of which there were four, were named *cas1-4*. Of note, Jansen et al. and Mojica et al. were the first to standardize the nomenclature of this field and settled upon CRISPR for the DNA arrays and Cas for the proteins encoded by the flanking genes. Following the first four *cas* genes, a large number of other genes were identified and grouped into putative functional subtypes based on sequence homology (Haft et al. 2005; Makarova et al. 2006). The function of these *cas* genes was confirmed after it was demonstrated that the proteins they encode process RNA transcribed from CRISPR loci (crRNA) and form a large multi-protein complex with the crRNA, that silences foreign DNA (Brouns et al. 2008). The crRNA functions like a guide and binds complementary sequences, while the Cas proteins catalyze the degradation of foreign nucleic acid.

With these results, it is apparent that the CRISPR system must have a mechanism to acquire new spacers, generate crRNAs and then use Cas proteins to degrade foreign nucleic acid. As such, the CRISPR immune response has been divided into three distinct stages: (i) adaption, (ii) crRNA biogenesis, and (iii) interference (Makarova et al. 2011). In addition, the large number of *cas* genes flanking CRISPR loci have been divided into functional groups based upon sequence homology and biochemical function.

Structure of CRISPR loci

CRISPR loci consist of short repeating DNA sequences interspaced by variable invader sequences called spacers (Figure 1.2). The size of the repeats ranges from 24-47 base pairs and a large subset of them are palindromic (Grissa et al. 2007). When transcribed, these palindromic sequences can form secondary structures, such as hairpins (Mojica et al. 2005). The variable spacers range in size from 20-58 base pairs and are typically derived from extra-chromosomal sources, yet it has been shown that a small subset of spacers are derived from the host chromosome (Mojica et al. 2000; Stern et al. 2010). The longest CRISPR loci was found to contain 374 repeat-spacer sequences, and a single prokaryotic chromosome can carry between 1-18 different CRISPR loci (Marraffini & Sontheimer 2010a; Pourcel et al. 2005). The sequence of the repeats at different CRISPR loci for a given organism is typically very similar. Directly upstream to a CRISPR array is the leader sequence, which is A-T rich. The leader sequence also consists of promoter elements that control transcription of the CRISPR loci. Acquisition of new spacers always occurs near to the leader region. This results in polarity, whereby the most recently incorporated spacers are proximal to the leader sequence and older spacers are distal (Pourcel et al. 2005). As mentioned above, there is a finite amount of spacers that can be incorporated and those oldest are lost. Additionally, because of the polarity a CRISPR locus can serve as a chronological history of the infections a particular host has been subjected to.

The different CRISPR systems

Based on the overall sequence conservation of *cas* genes, the CRISPR system has been divided into five main types- I, II, III, IV, and V (Figure 1.3) (Makarova et al. 2015). This classification is the latest on the state of CRISPR organization, which has undergone several updates as new genes are discovered (Haft et al. 2005; Makarova et al. 2006; Makarova et al. 2011). Due to the extreme diversity of *cas* genes and the difficulty of generating simple phylogenetic classifications because of the dynamic nature of prokaryotic genomes, we can expect further updates to the organization of CRISPR systems in the future. I will focus on types I, II, and III, which are better characterized, to give a survey of their activity through the three stages of the CRISPR immune response (Figure 1.4).

The CRISPR type I system is the most prevalent and is identified in prokaryotic genomes through the presence of its signature gene *cas3* (Makarova et al. 2011). The Cas3 protein is involved in the interference step and it has been demonstrated to degrade foreign DNA (Brouns et al. 2008; Jore et al. 2011; Semenova et al. 2011). Cas3 usually consists of an N-terminal nuclease domain and a C-terminal helicase domain, although the two domains can sometime exist as two separate genes. The nuclease domain and specifically its histidine-aspartate motif have been shown to be responsible for cleavage of ssDNA (Mulepati & Bailey 2011; Sinkunas et al. 2011). The type I system can be

further divided into seven subtypes (I-A to I-F and IU), each with different sets of Cas proteins that form large multi-protein complexes with the crRNA. These complexes help process the crRNA and then bind to dsDNA that matches the crRNA it is carrying. Upon binding the target dsDNA, the complexes form R-loops and Cas3 is recruited to degrade the exposed ssDNA.

The CRISPR type II system is only found in bacteria and its signature gene is *cas9* (Makarova et al. 2011). The type II system, unlike types I or III, requires only one protein, Cas9, to identify and degrade foreign sequences. Cas9 contains two well-conserved domains, HNH and RuvC, both of which are involved in processing crRNA and cleaving target DNA. The type II system is unique in that, it requires a trans-activating crRNA (tracrRNA) and endogenous RNase III to process crRNA (Deltcheva et al. 2011). The type II effector complex, which includes Cas9, crRNA, and tracrRNA, can perform dsDNA cleavage. Using the crRNA as a guide, Cas9 binds to dsDNA and cleaves the strand complementary to the crRNA with its HNH nuclease domain and cleaves the non-complementary strand with its RuvC nuclease domain (Osawa et al. 2015). The type II system is further divided into three subtypes (II-A to II-C).

The CRISPR type III system is more abundant in archaea than bacteria and is composed of four subtypes (III-A to III-D) (Staals et al. 2014; Tamulaitis et al. 2014). The signature gene is *cas10* and it is involved in the interference step. Type III systems are similar to type I, in that they form large multi-protein complexes that bind crRNA and

protein crystal diffraction studies show that they are structurally comparable (Deng et al. 2013; Goldberg et al. 2014; Samai et al. 2015). Unlike the type I complexes, the type III complexes are not involved in the crRNA-processing step although they may play a role in secondary processing (Zhang et al. 2012). A separate protein that is not part of the effector complex, Cas6, performs primary crRNA processing in the type III system (Samai et al. 2015). Additionally, the multi-protein complexes of type III systems are sufficient to cleave their targets and don't require the recruitment of additional factors, like Cas3 in the type I system (Taylor et al. 2015). The type III system has been demonstrated to cleave both DNA and RNA (Mulepati & Bailey 2011).

It is worth mentioning that a single prokaryote may carry various subtypes belonging to the various CRISPR types. In addition, different effector complexes can use the same spacers (Majumdar et al. 2015), although they will be processed differently based on the CRISPR system type (Carte et al. 2014).

The three stages of CRISPR

The CRISPR system has three stages: (i) adaption, (ii) crRNA biogenesis, and (iii) interference (Figure 1.5). The adaption stage is conserved among all CRISPR types, but the crRNA biogenesis and interference stages are quite different. In addition, there is evidence of coordination between the stages orchestrated by the various Cas proteins.

Adaption

The adaption stage provides prokaryotes with the genetic memory to mount an immune response when challenged by invasive genetic elements. The acquisition of new spacers has been demonstrated for different organisms containing different CRISPR arrays and subtypes, indicating a common mechanism between all organisms containing CRISPR (Makarova et al. 2011; Makarova et al. 2015). There are two methods to acquire spacers: naïve, when the invader is encountered for the first time, and primed, when there is a pre-existing record of the invader in the CRISPR loci. In both methods, the Cas1 and Cas2 proteins are key factors in spacer integration. This function has been confirmed by demonstrating that they are sufficient for integration to occur in the absence of all other Cas proteins and that they directly interact with CRISPR DNA (Jackson & Wiedenheft 2015). In addition, Cas1 and Cas2 are considered essential genes for CRISPR because they are ubiquitously found in all prokaryotes containing CRISPR. However, other Cas proteins have been implicated including Cas9, Csn2, and Cas4 (Mulepati & Bailey 2011). The role of these proteins and other Cas proteins in spacer integration is not directly known.

The invader sequence selected for acquisition into a CRISPR array is referred to as a protospacer. Protospacers are small (20-58 base pairs) relative to the entire genome of the invader to be surveyed. *In silico* analysis has shown that a 2-5 base pair sequence adjacent to the protospacers in the invader genome are conserved within a species

(Sinkunas et al. 2013). This sequence, called a protospacer adjacent motif (PAM), has been shown to be essential for spacer acquisition (Makarova et al. 2012). The PAM is also important for the interference stage to help find target sequences for degradation. In addition to the PAM, other DNA motifs that help select protospacers for integration have been identified including a 2 base pair sequence called acquisition affecting motif (AAM) and Chi sites which are enriched in high copy number invasive plasmids and viruses (Hatoum-Aslan et al. 2014).

During primed acquisition of invader sequences, effector complexes that are typically associated with the silencing stage, bind to foreign nucleic acid protospacers though the complementarity between the guide crRNA and the target may not be perfect. The effector complex will not be able to degrade the foreign nucleic acid because of the imperfect match in sequence but this process leads to the recruitment of other Cas proteins, resulting in an accelerated uptake of spacers from the invader. This is advantageous because multiple spacers to the same invader provide increased resistance and make it more difficult for the target to evolve escape mutants, as several sites would need to be changed simultaneously. The process of primed acquisition has been demonstrated in the type I-E system of *E. coli* (Deng et al. 2013).

Mechanistically to integrate a spacer, the first repeat sequence proximal to the leader sequence serves as a template for the synthesis of a new repeat and the spacer, cleaved by Cas1-Cas2, is integrated between them (Hale et al. 2009; Hatoum-Aslan et al.

2014). The palindromic nature of CRISPR repeats leads to their adopting a cruciform structure, which is important to determine the position and direction of spacer integration into the array (Hale et al. 2009; Zhang et al. 2012; Staals et al. 2013; Staals et al. 2014; Tamulaitis et al. 2014; Samai et al. 2015). Taken together, spacer integration is orchestrated by both the sequence and structure of CRISPR loci. Occasional deletion of spacers is required to limit the size of the CRISPR, but there is little knowledge of the mechanism or frequency of such events.

crRNA biogenesis

crRNA biogenesis involves the transcription CRISPR loci, which generates a long RNA that is then processed into smaller functional crRNAs that can guide effector complexes to their targets. The processed crRNAs are composed of a single spacer flanked by partial repeats. Expression of the CRISPR locus is driven by the leader sequence, which is A-T rich and consist of promoter elements (Samai et al. 2015). How CRISPR expression is induced is still not clear, but there is evidence that there are transcriptional activators (Makarova et al. 2012). In addition, envelop stress has been shown to induce expression, this is significant because when viruses bind to prokaryotic cells they also cause envelop stress (Deng et al. 2013; Goldberg et al. 2014; Samai et al. 2015).

It has been demonstrated that even when different CRISPR subtypes co-exist in an organism, they process their own crRNAs (Goldberg et al. 2014; Jackson & Wiedenheft 2015). The crRNA biogenesis stage is not conserved, and there are significant differences between type I, II and III. In the type I system, after a CRISPR locus is transcribed, the repeats form stable stem loops that are recognized and cleaved by a Cas6 family endoribonuclease (Hale et al. 2014). The products of this endoribonuclease reaction are 61 nucleotide long crRNAs that contain a unique spacer flanked by partial repeat segments. After processing, Cas6 remains bound to the crRNA and other type I Cas proteins bind to it forming a large multi-protein complex (Tamulaitis et al. 2014). The type I complex is known as the CRISPR-associated complex for antiviral defense (Cascade). In the type II system, a trans-encoded short RNA, termed trans-activation crRNA (tracrRNA), is required for crRNA processing (Samai et al. 2015). The tracrRNA has a 25-nucleotide segment that is almost always complementary to the repeat sequence in CRISPR loci, and the base pairing is required for the pre-cRNA and tracrRNA co-processing. The processing is done by the endogenous RNase III nuclease. The type III system utilizes a similar approach as the type I system. However, without the stem-loop feature in the repeats (Wang & Landick 1997), the endoribonuclease Cas6 binds to the first few bases from the 5' end of the repeat and cleaves at 8 nucleotides upstream from the spacer (Makarova et al. 2011). Unique to type III systems, the 3' end of the crRNA is further processed by an unknown mechanism at 6 nucleotide intervals and can generate mature crRNAs that contain variable repeat sequence on the 5' end or just spacer sequence (Jung et al. 2015). Unlike the type I system, in the type III systems the crRNA

is handed off to a multi-protein effector complex containing subtype specific Cas proteins and Cas6 is not part of the complex (Westra et al. 2012; Sinkunas et al. 2013; Mulepati & Bailey 2013).

Interference

The interference stage of the CRISPR III-B subtype is the primary subject of my thesis, yet I will briefly delve into what is known about type I and II systems. After the assembly of the effector complex composed of crRNA and Cas protein(s), the foreign genetic element to which the crRNA is complementary can be recognized and degraded. The Cas proteins catalyze the destruction of invader sequences through different mechanisms.

In the type I system, Cascade acts as a surveillance complex that will bind to protospacers to which its crRNA is complementary. Upon binding to a target protospacer, Cascade melts the double stranded target DNA and forms an R-loop with the complementary strand bound to the crRNA and the non-complementary strand is exposed (Hatoum-Aslan et al. 2014). This is believed to be a signal for Cas3 recruitment, which has Mg^{2+}-dependent ssDNA cleavage activity and an ATP-dependent helicase domain, to degrade the protospacer containing foreign DNA (Hatoum-Aslan et al. 2014; Ramia, Tang, et al. 2014b; Samai et al. 2015). Cascade binding is ATP-independent and is proposed to nucleate at the 5' end of the protospacer, where the seed sequence is, and

Watson-Crick base-pairing proceeds to the 3' end (Staals et al. 2013). Binding is dependent on PAM, and in the silencing stage PAM is the determining factor in self-vs-foreign targeting. CRISPR loci do not contain PAM sequences.

In the type II system Cas9, with its tracrRNA and crRNA, binds to protospacers complementary to the crRNA. Binding to the PAM is the first step in target interference for the type II system (Garrett et al. 2015; Hale et al. 2012). This 3-5 base pair sequence serves as an initial screen of the invasive DNA sequence, if a match is found subsequent base pairing to the crRNA ensues in a manner similar to Cascade. Cas9 contains two nuclease active sites, the HNH domain and the RuvC-like domain, that allow it to cleave dsDNA directly. The complementary strand of the target DNA is cleaved by HNH nuclease domain while the non-complementary strand is cleaved by RuvC-like domain (Deng et al. 2013).

In the type III system, multi-protein effector complexes target both DNA and RNA for degradation. Discrimination between self and non-self doesn't occur through a PAM sequence like in the type I and II systems, instead a 5' tag derived from the CRISPR repeat sequences on the crRNA biases foreign nucleic acid, with complementarity to the 5' tag indicating self. The type III systems bind and cleave single-stranded (ss) RNA (Samai et al. 2015), and transcriptionally active DNA (Peränen et al. 1996). The Type III systems can be further classified into subtypes and the Type III-A and Type III-B subtypes are the best characterized. Both subtypes contain a Cas10

family protein, Cmr2 in Type III-B and Csm1 in Type III-A systems (Huang & Szostak 1996). The III-A subtype has been shown to cleave ssRNA complementary to the crRNA at 6 nucleotide intervals by the Csm3 protein (Staals et al. 2014; Samai et al. 2015; Tamulaitis et al. 2014), but they have also been demonstrated to cleave DNA in a transcription-coupled manner. The III-B complex has also been demonstrated to cleave ssRNA at 6 nucleotide intervals by the Cmr4 protein (Benda et al. 2014; Ramia, Spilman, et al. 2014a). *In vivo* genetic experiments have implicated the III-B complex in targeting transcriptionally active DNA, but the mechanism of this activity and the identity of the DNA nuclease active site are unknown.

Thesis rationale

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their associated (Cas) proteins form the basis of an inheritable and adaptable RNA-guided immune system in prokaryotes. For a recent review see (van der Oost et al. 2014). CRISPR loci are composed of repeat sequences separated by variable spacer sequences that are derived from foreign genetic elements (Bolotin et al. 2005; Mojica et al. 2005; Pourcel et al. 2005; Barrangou et al. 2007). Transcripts from CRISPR loci are processed to generate short CRISPR RNAs (crRNA) (Brouns et al. 2008; Carte et al. 2008; Deltcheva et al. 2011). The crRNA are then incorporated into effector complexes that identify and destroy invading nucleic acid that is complementary to the guide region of the crRNA (Brouns et al. 2008; Hale et al. 2009; Jinek et al. 2012; Gasiunas et al. 2012).

CRISPR-Cas systems are organized into three types (Type I, II and III) and at least ten subtypes (Makarova et al. 2011). Each type has a distinct mechanism for processing CRISPR transcripts and for target destruction (Makarova et al. 2011) and are distinguished by a signature protein, Cas3 in Type I, Cas9 in Type II and Cas10 in Type III. The effector complexes in Type I systems recognize their DNA targets and then recruit the Cas3 protein to degrade the invading DNA using its Histidine-Aspartate (HD) nuclease domain (Brouns et al. 2008; Westra et al. 2012; Mulepati & Bailey 2013; Sinkunas et al. 2013). Type II systems also target DNA but recognition and cleavage are mediated by the same effector complex, Cas9 (Jinek et al. 2012; Gasiunas et al. 2012).

The Type III systems bind and cleave single-stranded (ss) RNA (Hale et al. 2009; Zhang et al. 2012; Zebec et al. 2014; Staals et al. 2013; Staals et al. 2014; Tamulaitis et al. 2014; Samai et al. 2015; Hale et al. 2012) and transcriptionally active DNA (Goldberg et al. 2014; Samai et al. 2015; Deng et al. 2013). The Type III systems can be further classified into Type III-A and Type III-B subtypes. Both subtypes contain a Cas10 family protein, Cmr2 in Type III-B and Csm1 in Type III-A systems (Makarova et al. 2011). Despite mechanistic differences, the Type I and Type III effector complexes share similar structures and likely evolved from a common ancestor (Jackson & Wiedenheft 2015).

The Cmr complexes from the Type III-B systems of *Pyrococcus furiosus, Thermus thermophilus* and *Sulfolobus solfataricus* have all been characterized *in vitro* (Hale et al. 2009; Staals et al. 2013; Zhang et al. 2012). The Cmr effector complex consists of six proteins (Cmr1-6) and a crRNA. The crRNA is comprised of eight nucleotides of repeat sequence at its 5' end (the 5' handle) and typically 30-40 nucleotides of invader-derived guide sequence at its 3' end (Carte et al. 2008; Hale et al. 2009; Staals et al. 2013). Multiple Cmr4 and Cmr5 subunits form a helical filament around the crRNA, which is capped at its 5' end by Cmr2 and Cmr3 and at its 3' end by Cmr6 and Cmr1 (Spilman et al. 2013; Staals et al. 2013; Taylor et al. 2015; Benda et al. 2014; Osawa et al. 2015). Cmr4 serves as a catalytic subunit (Benda et al. 2014; Ramia, Spilman, et al. 2014a; Zhu & Ye 2015) cleaving paired ssRNA targets at regular 6 nt intervals (Hale et al. 2014; Staals et al. 2013). Cmr2 contains a predicted HD nuclease domain (Makarova et al. 2011), although this domain is dispensable for RNA target

cleavage by the Cmr complex (Cocozaki et al. 2012; Staals et al. 2014). The Csm complexes from the Type III-A systems have also been characterized *in vitro*. They have a similar structure to the Cmr complex (Staals et al. 2014; Rouillon et al. 2013) and likewise cleave complementary RNA at multiple sites (Samai et al. 2015; Staals et al. 2014; Tamulaitis et al. 2014) but have also been shown to cleave DNA in a transcription-coupled manner (Samai et al. 2015).

In cells, genetic experiments have implicated the Cmr complex in targeting transcriptionally active DNA (Deng et al. 2013) but the mechanism of this activity and the identity of the DNA nuclease active site are unknown. We therefore undertook a biochemical analysis of the Cmr complex to determine if and how it can cleave DNA.

Figures

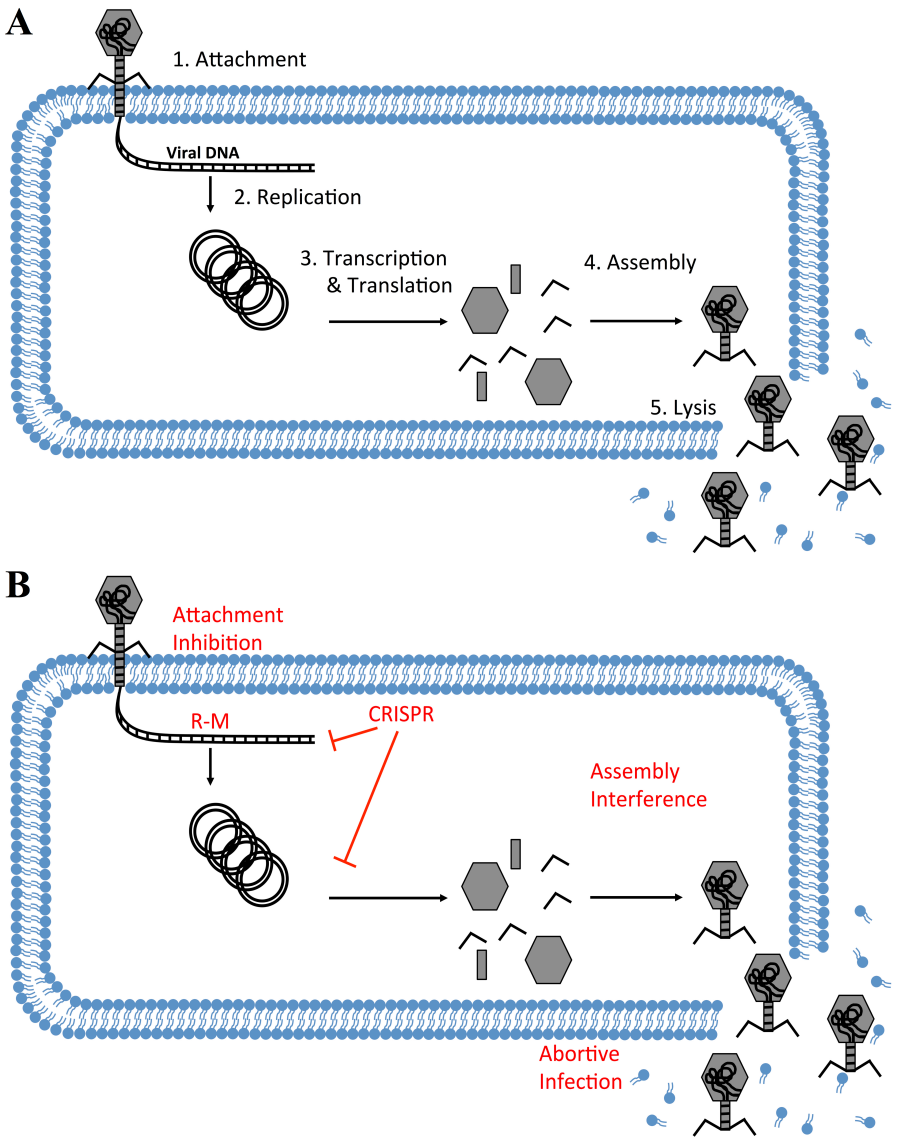


Figure 1.1 An overview of the viral life cycle and bacterial defense strategies. (A) The life cycle stages of a lytic virus. (B) Bacterial strategies to combat different stages of the viral life cycle. For simplicity, the outer membrane (gram-negative bacteria) and cell wall are not shown.

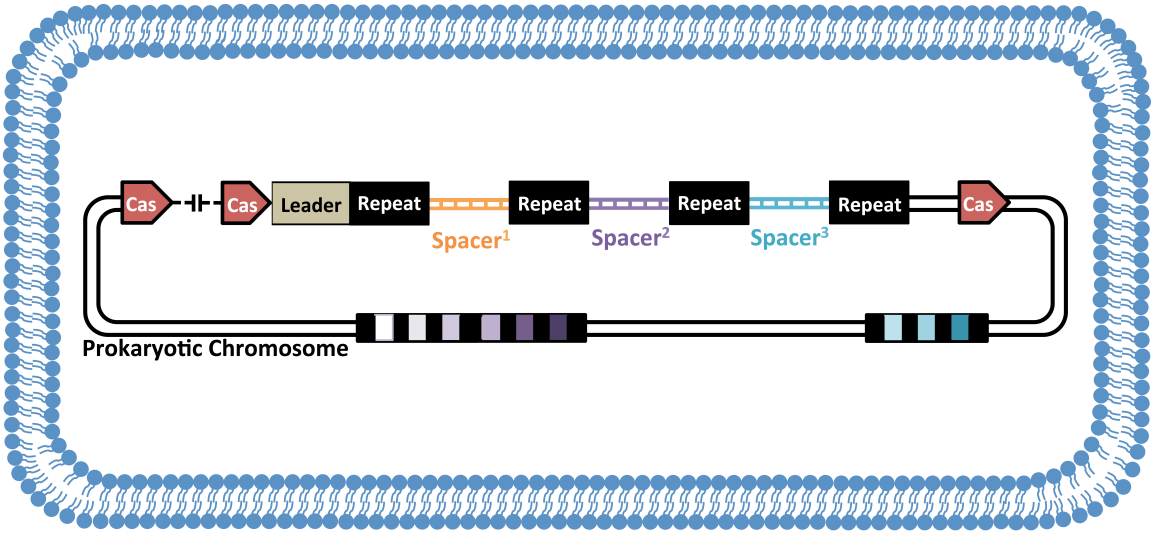


Figure 1.2 Representative CRISPR loci in a Prokaryotic Chromosome. The repeat sequences are shown in black, the spacers are colored and the *cas* genes are red. The spacers are numbered from newest (Spacer[1]) to oldest (Spacer[3]) integrated sequence. The leader sequence (tan colored), which regulates CRISPR transcription, is proximal to Spacer[1].

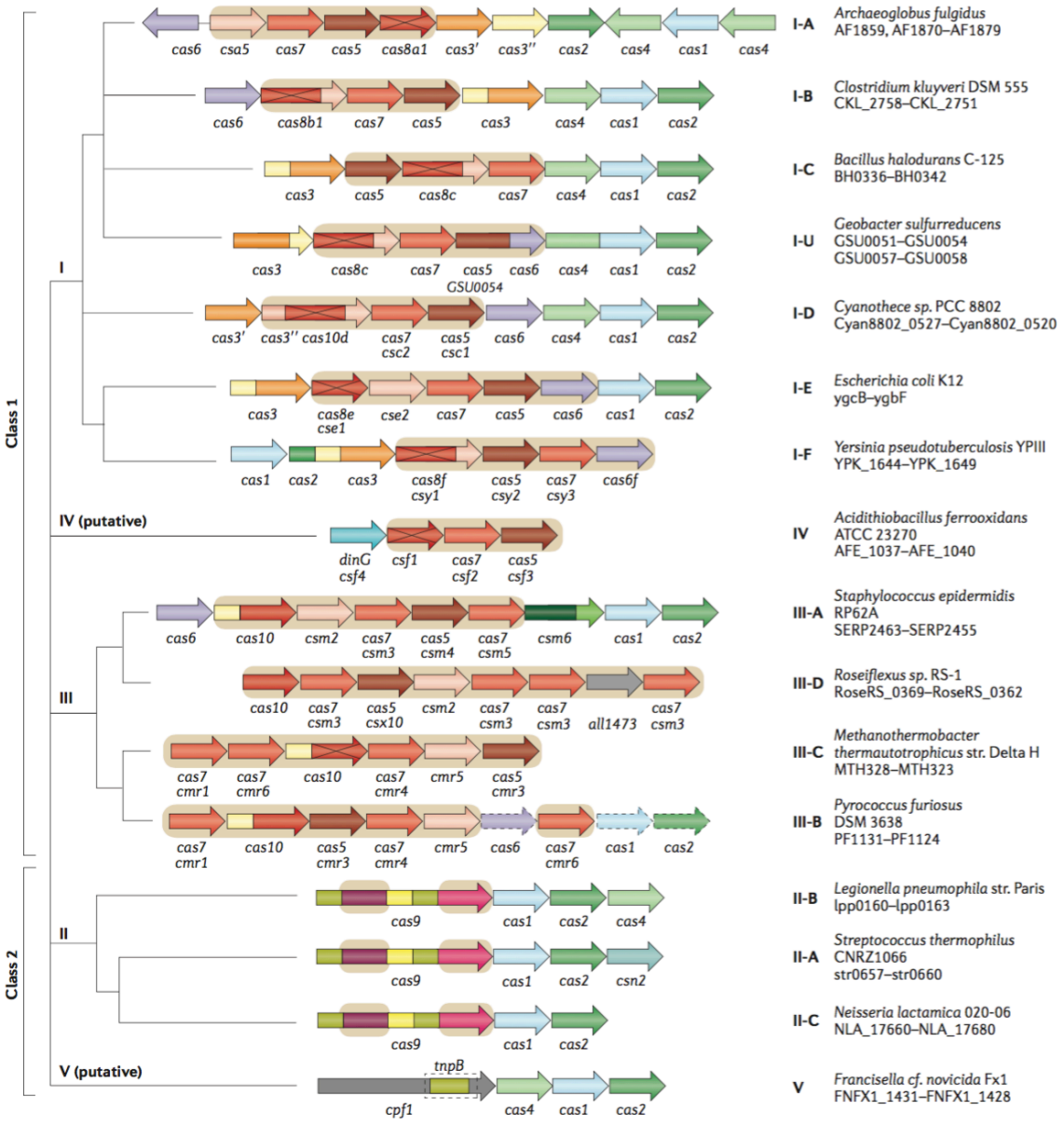


Figure 1.3 Diversity of Cas proteins (Makarova et al. 2015). There are 5 types (I-V), which are further subdivided into different subtypes.

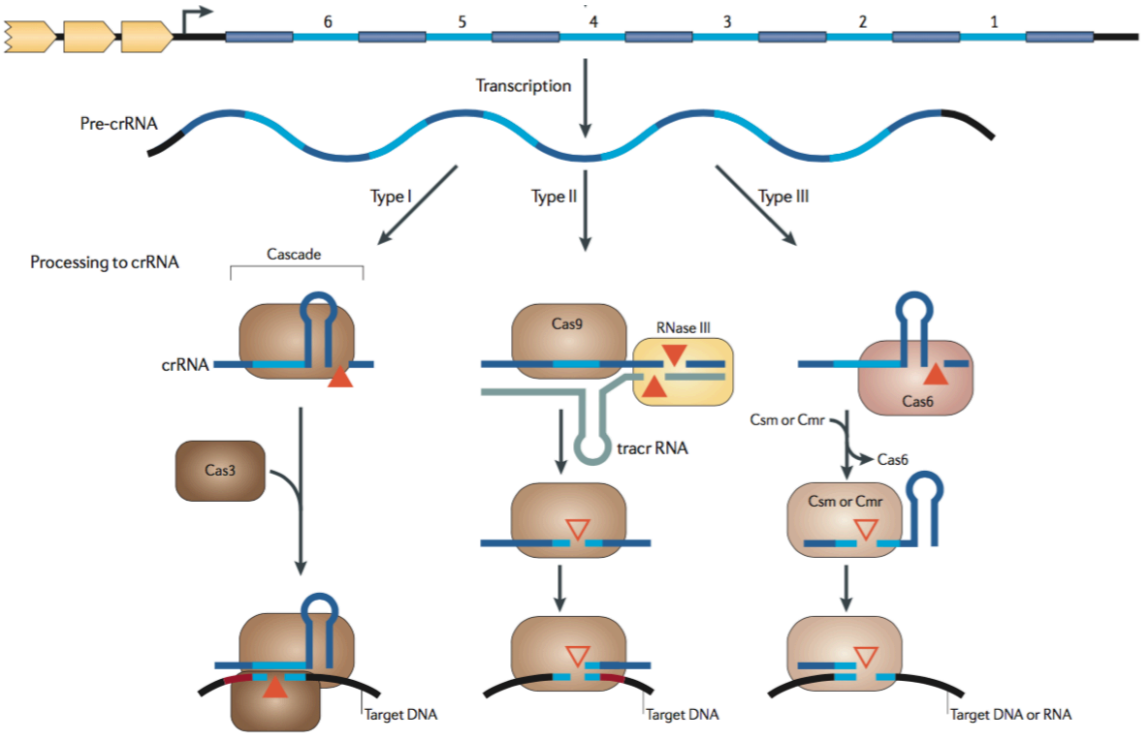


Figure 1.4 Type I, II, and III CRISPR systems (Makarova et al. 2011). The signature proteins of the type I, II, and III systems are Cas3, Cas9, and Cas10, respectively. The systems are classified based on the phylogenetic relationships of the *cas* genes.

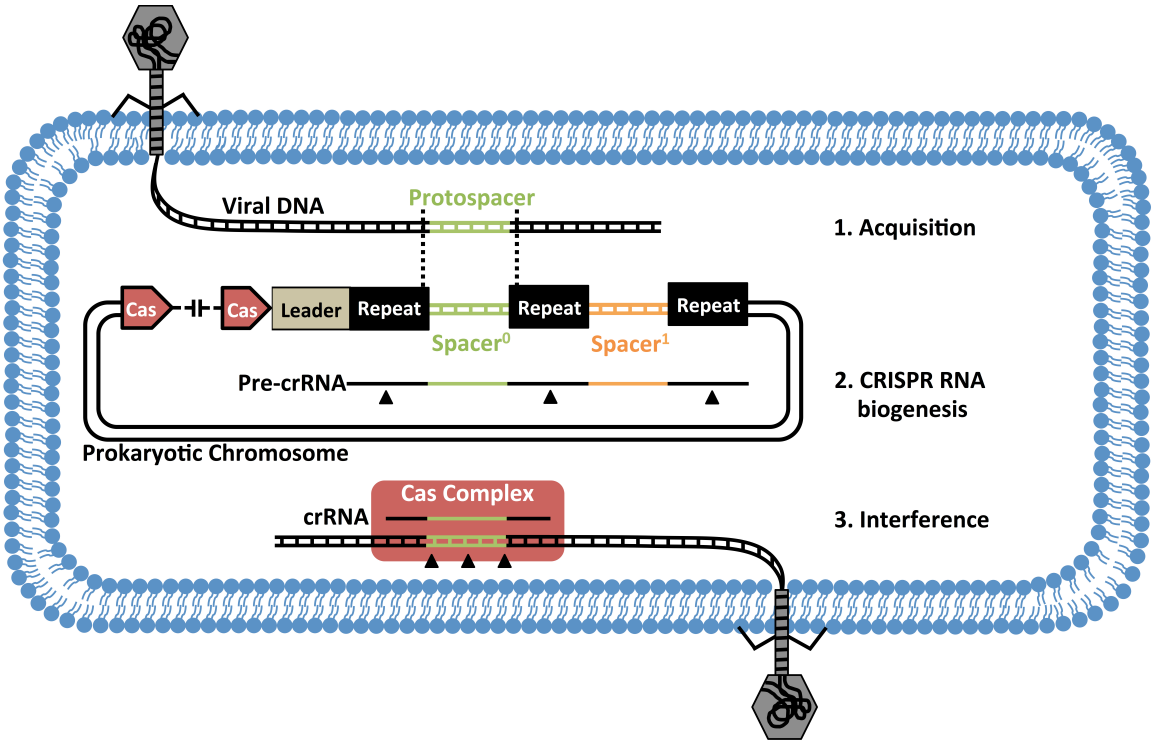


Figure 1.5 The CRISPR immune system has three distinct stages. First, spacers complementary to the foreign protospacer are incorporated into the host genome. Second, subsequent attacks elicit transcription at CRISPR loci and the generation of CRISPR RNA through processing by Cas proteins. Third, CRISPR RNA is loaded onto Cas effector complexes, which target and cleave nucleic acid complementary to the CRISPR RNA.

CHAPTER 2

MAKING THE TYPE III-B COMPLEX

Reconstitution of the Cmr complex from *Thermotoga maritima*

The genome of *Thermotoga maritima* MSB8 contains three Cas modules (one each of types I-B, III-A and III-B) and eight CRISPR loci (Figure 2.1). Each CRISPR loci contains the same, or very similar, repeat sequence, suggesting that the three Cas modules can likely use crRNA processed from any of the CRISPR loci. The Type III-B module encodes six proteins (Cmr1-6). To determine if these six proteins can assemble into a Cmr complex we cloned, expressed and purified each *Thermotoga maritima* Cmr protein in an *Escherichia coli* expression system (Figure 2.2A). The recombinant proteins were mixed and the mixture purified over a nickel column (Cmr6 was N-terminally hexa-histidine tagged, all of the other proteins were untagged) in either the presence or absence of a synthetic crRNA8.3 (which corresponds to the third crRNA encoded by CRISPR locus 8). In the absence of crRNA all but one of the Cmr proteins, Cmr1, purified as a complex, whereas in the presence of crRNA, all six Cmr proteins purified as a complex (Figure 2.2B). Weak association between Cmr1 and the Cmr complex has been noted in both *T. thermophilus* (Staals et al. 2013)and *P. furiosus* (Benda et al. 2014).

We next asked if the organization of the *Thermotoga maritima* Cmr complex is similar to Cmr complexes from other organisms. To test this, we performed a series of pull-down experiments with combinations of *Thermotoga maritima* Cmr proteins. The results of these experiments are presented in Figure 2.3. The interactions between subunits observed in the *Thermotoga maritima* Cmr complex agrees with the interactions

observed between subunits of other Cmr complexes by electron microscopy, cross-linking and X-ray crystallography (Spilman et al. 2013; Staals et al. 2013; Taylor et al. 2015; Benda et al. 2014; Osawa et al. 2015), suggesting that the *Thermotoga maritima* Cmr complex has the same subunit architecture as other Cmr complexes.

Materials and methods

Purification of recombinant Cmr proteins

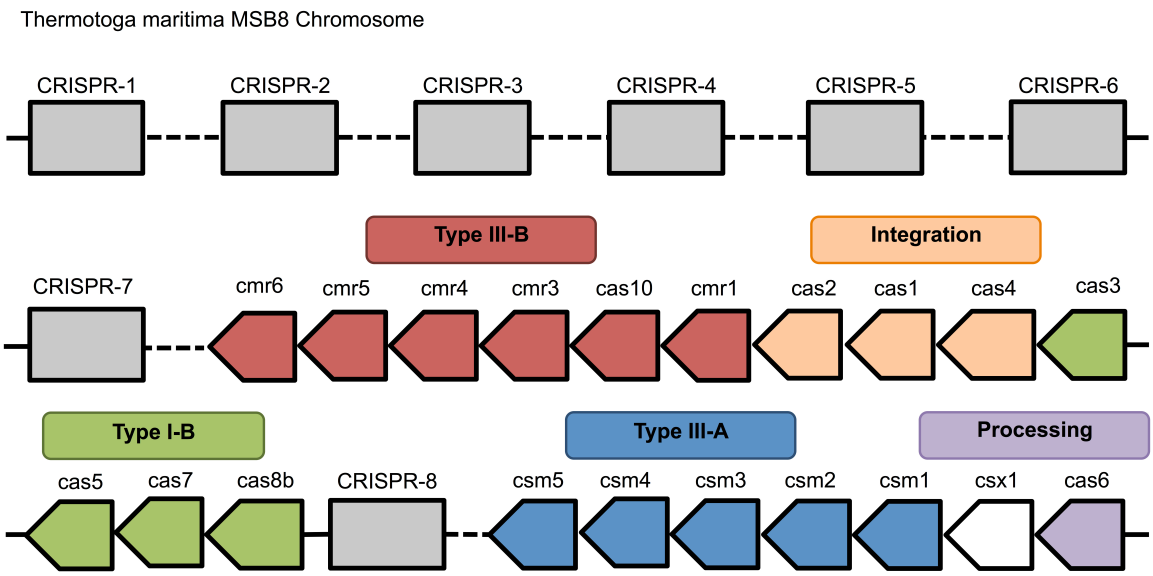
The genes encoding Cmr1-6 were individually PCR amplified from *Tma* genomic DNA and cloned into pET derived vectors. The vectors, pHAT4 and pHAT2, contain an amino-terminal hexahistidine tag that is cleavable or not cleavable by tobacco etch virus protease (TEV), respectively. The vectors were transformed into T7Express cells (New England BioLabs) and grown to an OD_{600} of 0.3 in Luria-Bertani Medium, followed by induction with 0.2mM isopropyl-β-D-thiogalactopyranoside (IPTG) and overnight expression at room temperature. The cells were pelleted at 4,500 rpm, resuspended in lysis buffer (1 M KCl, 20 mM Tris-HCl at pH 8.0, 10 mM Imidazole and 1 mM tris(2-carboxyethyl)phosphine (TCEP)) and disrupted using a French Press. The resulting sample was centrifuged at 18,000 rpm for 45 minutes at 4°C. The histidine tagged Cmr proteins were isolated from the supernatant using a 5mL IMAC column (Bio-Rad) charged with nickel sulfate and equilibrated with lysis buffer. After washing with at least 100mL lysis buffer, the proteins were eluted with 250mM Imidazole containing lysis buffer and injected onto a HiLoad 26/60 S200 size exclusion column (GE Healthcare) equilibrated with gel filtration buffer (350 mM KCl, 20 mM Tris-HCl at pH 8.0 and 1 mM TCEP). The above steps were carried out with Cmr proteins cloned into the pHAT2 vector. When cloned into the pHAT4 vector histidine tag removal was accomplished by elution off of IMAC as above, followed by buffer exchange with a HiPrep 26/10

desalting column (GE Healthcare) equilibrated with gel filtration buffer and incubation with TEV protease overnight at 4°C. The sample is re-applied to the IMAC column to remove the cleaved histidine tag and histidine tag containing TEV protease, the flow-through is collected and injected onto a HiLoad 26/60 S200 size exclusion column equilibrated with gel filtration buffer. Fractions collected from the size exclusion column were analyzed by SDS-PAGE. The Cmr2 and Cmr4 mutants were created by PCR-based mutagenesis. The expression vector was amplified using phosphorylated mutagenic primers followed by self-ligation. All mutant proteins were expressed and purified using the same protocol as the respective wild-type protein.

Pull-down assay

Purified Cmr proteins that were tagged (bait) and untagged (prey) were incubated in 100μL binding buffer (100 mM KCl, 20 mM Tris-HCl at pH8.0, 10 mM Imidazole, 1 mM TCEP) for 30 min at 37°C. 20μL of nickel sulfate charged IMAC resin (BioRad) equilibrated in binding buffer is added to the Cmr protein sample and incubated for 1 hr at 4°C on a rotating plate. The resin was washed three times with 0.5 mL binding buffer and further incubated with elution buffer (binding buffer containing 250 mM Imidazole) for 30 min at 4°C. Proteins eluted from the beads were analyzed on 4-20% Mini-Protean TGX precast gels (BioRad) stained with Coomassie.

Figures



Locus	Genomic position	Repeat sequence	# Spacers
1	412-3,015	GTTTCAATAATTCCTTAGAGGTATGGAAAC	40
2	369,170-369,731	GTTTCAATAATTCCTTAGAGGTATGGAAAC	8
3	396,568-397,133	GTTTCAATAATTCCTTAGAGGTATGGAAAC	8
4	409,295-410,923	GTTTCAATAATTCCTTAGAGGTATGGAAAC	24
5	1,428,083-1,428,243	GTTTCAATAATTCCTTAGAGGTATGGAAAC	2
6	1,519,912-1,520,474	GTTTCAATACTTCCTTAGAGGTATGGAAAC	8
7	1,631,790-1,632,016	GTTTCAATACTTCCTTAGAGGTATGGAAAC	3
8	1,779,810-1,780,701	GTTTCAATAATTCCTTAGAGGTATGGAAAC	13

Figure 2.1 Schematic representation of the CRISPR arrays and cas genes in the genome of *Thermotoga maritima* MSB8. CRISPR arrays are grey, Type III-B genes are red, Type I-B are green and Type III-A are blue. Genes involved in Integration and Processing are orange and purple, respectively. The repeat sequence of each array is shown, with the 5' handle resulting after processing underlined.

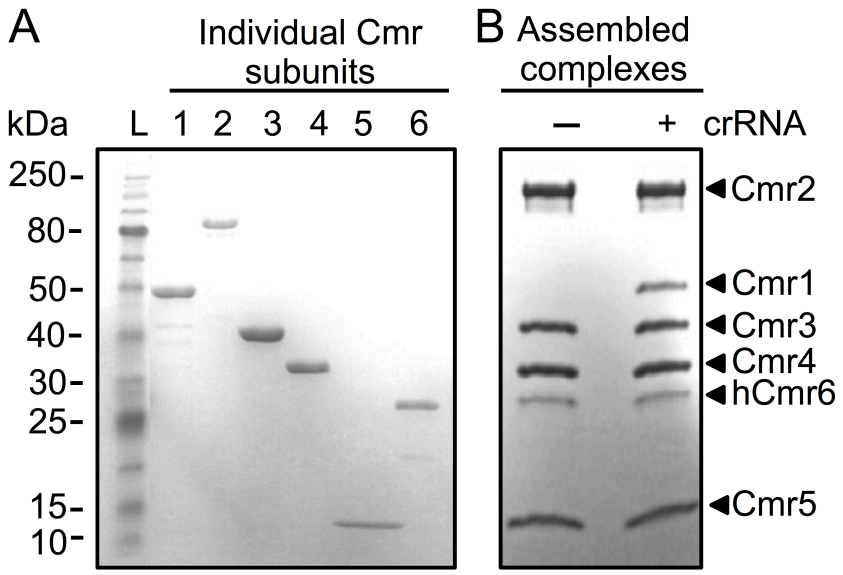


Figure 2.2 Assembly of the *Thermotoga maritima* Cmr complex. SDS-polyacrylamide gels, stained with Coomassie blue, of individually purified Cmr subunits (A) and affinity-purified Cmr complex (B) with and without crRNA8.3.

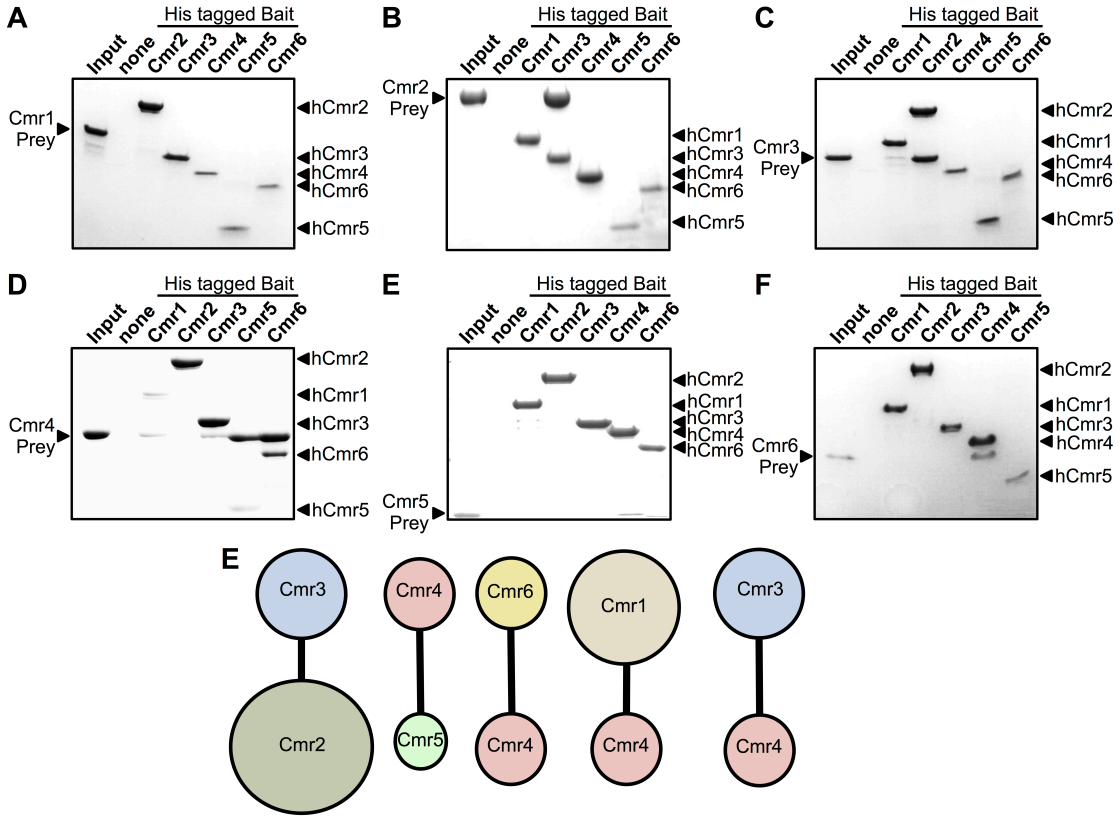


Figure 2.3 Pairwise interactions between Cmr subunits. (A-F) Pairwise pull-downs with each of the Cmr subunits as untagged prey against N-terminal His-tagged bait. (G) Schematic summary of the interactions observed in (A-F).

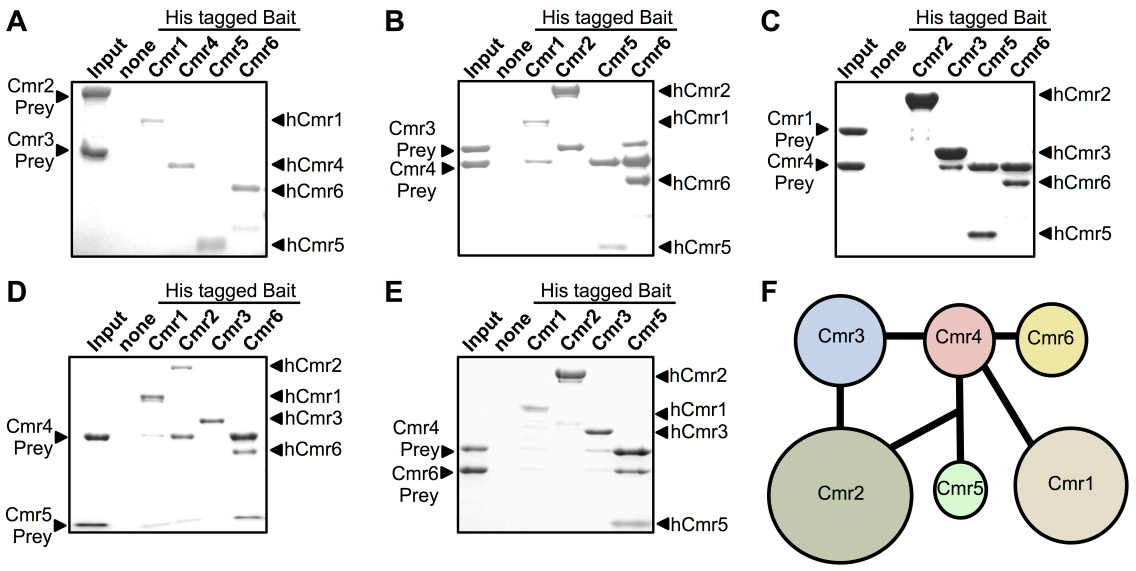


Figure 2.4 Interactions within the Cmr complex. (A-E) Pull-down experiments using the binary complexes identified in Figure 2.3 as untagged prey against N-terminal His-tagged bait. (F) Schematic summarizing all of the observed interactions.

CHAPTER 3

RNA DEGRADATION BY THE CRISPR TYPE III-B COMPLEX

The *Thermotoga maritima* Cmr complex cleaves ssRNA

In vitro the Cmr complexes from *P. furiosus*, *S. solfataricus* and *T. thermophilus* cleave ssRNA that is complementary to the crRNA (Hale et al. 2009; Zhang et al. 2012; Staals et al. 2013). To determine the substrate specificity of the *Thermotoga maritima* Cmr complex, we incubated the complex with crRNA8.3 and a selection of nucleic acid targets that were radiolabeled at their 5' ends (5'-labeled). Reactions were incubated for 10 minutes at 80°C, the optimal growth temperature of *Thermotoga maritima* (Huber et al. 1986), and the products were analyzed by denaturing PAGE followed by autoradiography. Neither non-complementary ssRNA, complementary ssDNA nor double-stranded (ds) RNA were cleaved (Figure 3.1A). Only complementary ssRNA (the ssRNA8.3 target) was cleaved, with the cleavage site located 14 nucleotides from the 3' end of the paired crRNA (Figure 3.1). Cleavage appeared to be sequence-independent as a ssRNA target complementary to a crRNA with a different sequence (crRNA8.4 and ssRNA8.4) also resulted in the same 14-nt cleavage product (Figure 3.1A), in agreement with results from the *P. furiosus* Cmr complex (Hale et al. 2012), and like other Type III complexes was dependent on the presence of magnesium or manganese ions (Figure 3.2) (Hale et al. 2009; Staals et al. 2014; Staals et al. 2013; Zhang et al. 2012; Samai et al. 2015; Tamulaitis et al. 2014). Further characterization of the cleavage activity revealed that the Cmr complex cleaves the target RNA on the 5' side of the phosphodiester bond like the *P. furiosus* Cmr complex (Hale et al. 2009). The cleavage product generated by the complex is not a substrate for polyadenylation (Figure 3.3), indicating the presence of

a 3’ phosphate (or 2’, 3’ cyclic phosphate) end. Target cleavage also required that the 5’ handle of the crRNA was intact and had the correct repeat-derived sequence (Figure 3.1B).

Other Cmr complexes can cleave their RNA targets at up to five sites that are separated by 6 nt intervals (Staals et al. 2013; Hale et al. 2014). To determine the frequency and spacing of target cleavage by the *Thermotoga maritima* Cmr complex, we monitored cleavage of a 5’-labeled ssRNA8.3 target as a function of time. The results show the target was cleaved at four sites (sites 1, 2, 3 and 4), each separated by 6 nt, producing 32, 26, 20 and 14-nt products (Figure 3.1C). Cleavage at site 1 is minimal compared to the other three sites, which was also observed with the *P. furiosus* and *T. thermophilus* Cmr complexes (Staals et al. 2013; Hale et al. 2014). The appearance and disappearance of the cleavage products over time suggest that the target is cleaved sequentially, starting predominantly at site 2 and then proceeding to site 3 then to site 4 (Figure 3.1E). After ~5 minutes the target is almost completely cleaved to a single prominent 14-nt product, which is much faster than observed with *P. furiosus* and *T. thermophilus* Cmr complexes (Hale et al. 2009; Staals et al. 2013). To confirm cleavage is sequential we repeated this experiment but used a 3’-labeled target. In this experiment we observed a single prominent 12-nt product (Figure 3.1D), corresponding to cleavage at site 2 (note that the 3’-end labeling reaction added one additional nucleotide to the 3’-end of the RNA target), suggesting that cleavage primarily begins at site 2.

To gain further insights into the determinants for RNA cleavage we tested activity on RNA targets that would result in base-pairing-disrupting mismatches when paired with crRNA8.3. RNA targets that result in mismatches at position 20 and/or position 21, which flank the scissile phosphate at site 3, do not inhibit cleavage of the target at any site. In fact, cleavage of these mismatched targets appears to be enhanced relative to the fully complementary target, as with these targets we observe an increased proportion of the site 4 product relative to the fully complementary target (experiments 1-4 in Figure 3.4A). Thus, perfect complementarity is not required at site 3, suggesting that RNA cleavage by the Cmr complex, like the related Csm complexes (Staals et al. 2014; Tamulaitis et al. 2014), is not strictly dependent on complete complementarity between the crRNA and RNA target.

It has been shown that the 2'-hydroxyl group adjacent to the scissile phosphate is crucial for the RNA cleavage activity of a chimeric Cmr complex reconstituted from *Archaeoglobus fulgidus* Cmr1-3 and *P. furiosus* Cmr4-6 (Osawa et al. 2015). We confirmed the importance of this hydroxyl group to cleavage by the *Thermotoga maritima* Cmr complex. An ssRNA8.3 target containing a 2'-deoxyribose at position 20 was not cleaved at site 3 (experiment 5 Figure 3.4A) but in a control experiment a target with a 2'-deoxyribose at position 21 was cleaved at site 3 (experiment 6 in Figure 3.4A). Furthermore, when any of the four cleavage sites were blocked with a 2'-deoxyribose we observed no cleavage at that site (experiments 7 through 11 in Figure 3.4B). A modified ssRNA8.3 target (ssRNA8.3*) that contained 2'-deoxyribose base substitutions at each of

the four cut sites (positions 14, 20, 26 and 32) was not cleaved by the Cmr complex (experiment 12 in Figure 3.4B). To confirm that the 2'-deoxyribose modifications in ssRNA8.3* do not interfere with binding to the Cmr complex we performed competition experiments. Thus, we monitored cleavage of the 5'-labeled ssRNA8.3 target using either unlabeled ssRNA8.3* as a specific competitor or ssRNA8.4 as an unspecific competitor. Only the unlabeled crRNA8.3* inhibited cleavage of the target (lane 3 in Figure 3.4C) thereby providing support for the Cmr complex binding to crRNA8.3*.

Materials and methods

Preparation of labeled oligonucleotides

RNA oligonucleotides were purchased from Sigma Aldrich (Table 1.1). They were 5' radiolabeled with T4 polynucleotide kinase (New England BioLabs) and γ-^{32}P ATP (Perkin Elmer) in 1x T4 polynucleotide kinase buffer at 37°C for 30 min. The substrates were resolved on a denaturing polyacrylamide gel, visualized by autoradiography, excised from the gel and placed in a 0.5mL solution of 0.3M sodium acetate overnight at 4°C followed by ethanol precipitation and resuspension in RNA Storage Solution (Ambion) for RNA. To generate dsRNA targets, labeled oligonucleotide with twice molar excess of complementary unlabeled oligonucleotide was incubated at 95°C for 10 min, followed by slow cooling to room temperature. Non-denaturing PAGE confirmed complete annealing.

Radiolabeling of the 3' end of ssRNA was performed as previously described (Huang & Szostak 1996). Briefly, ssRNA was annealed to a short ssDNA oligonucleotide, complementary to the 3'-end of ssRNA, generating a TG overhang. The duplex was then incubated with 3'-5' exonuclease deficient Klenow Fragment (New England BioLabs) and α-^{32}P dATP (Perkin Elmer) in 1x Buffer 2 (New England Biolabs) at 37°C for 2 hr. Following the reaction the ssDNA was removed by denaturing PAGE.

RNA cleavage assays

Cmr2-6 was formed by pull-down (see Chapter 2 Materials and Methods). Unless otherwise indicated, 250 nM Cmr2-6, 250 nM Cmr1, and 100 nM crRNA were incubated at 80°C for 30 min, in reaction buffer (100 mM KCl, 20 mM HEPES at pH 7, 1 mM $MnCl_2$, 0.5 mM ATP and 1 mM TCEP). The reaction (80°C) was initiated by addition of 1 nM (unless otherwise indicated) radiolabeled target. The reaction was stopped at 10 min or the indicated time point with the addition of quencher dye (90% formamide, 2.5% glycerol, 0.01% SDS, 0.01% bromophenol blue, 0.01% xylene cyanol and 1mM EDTA) and heated at 95°C for 10 min. Radiolabeled RNA ladders were generated with RNase T1 (New England Biolabs) or alkaline hydrolysis buffer (50 mM sodium carbonate pH 9.2 and 1 mM EDTA). The samples are then run on 20% polyacrylamide denaturing gels and visualized by phosphorimaging.

For analysis of the chemical ends of the cleavage products, cleavage reactions were performed using 5'-end labeled target as described above. The resulting RNA products were isolated by PCI extraction and ethanol precipitation, and subject to polyadenylation by incubation with 5 U *E. coli* polyA polymerase (NEB) for 15 min at 37°C as described by the manufacturer. The reaction was stopped by PCI extraction, followed by ethanol precipitation. The resulting products were analyzed on 20% polyacrylamide, TBE 7M Urea gels as described above.

Figures

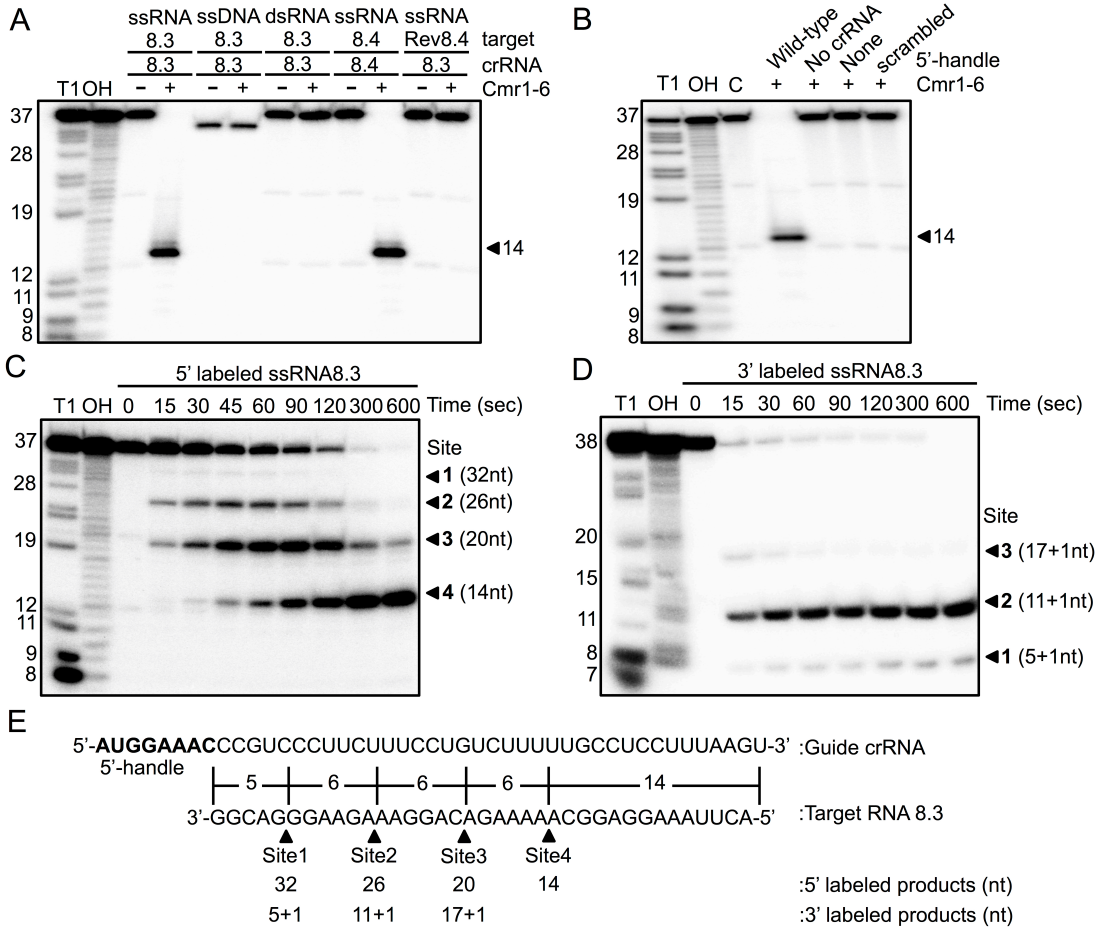


Figure 3.1. The *Thermotoga maritima* Cmr complex cleaves ssRNA. (A) The indicated 5'-labeled substrates were incubated with (+) or without (-) the Cmr complex and the products resolved on a denaturing polyacrylamide gel. (B) Requirement of the crRNA and determinant of its repeat derived 5'-handle. (C, D) 5'- and 3'-labeled ssRNA8.3 targets were incubated with the Cmr complex for the indicated times and resolved on

denaturing polyacrylamide gels. In all panels, T1 and OH denote the T1 and hydroxide ladders, respectively. (E) Schematic depicting the cleavage sites on the ssRNA8.3 target in relation to crRNA8.3. The sizes of the cleavage products are indicated. Note that the 3'-end labeling reaction added one additional nucleotide to the 3'-end of the RNA target.

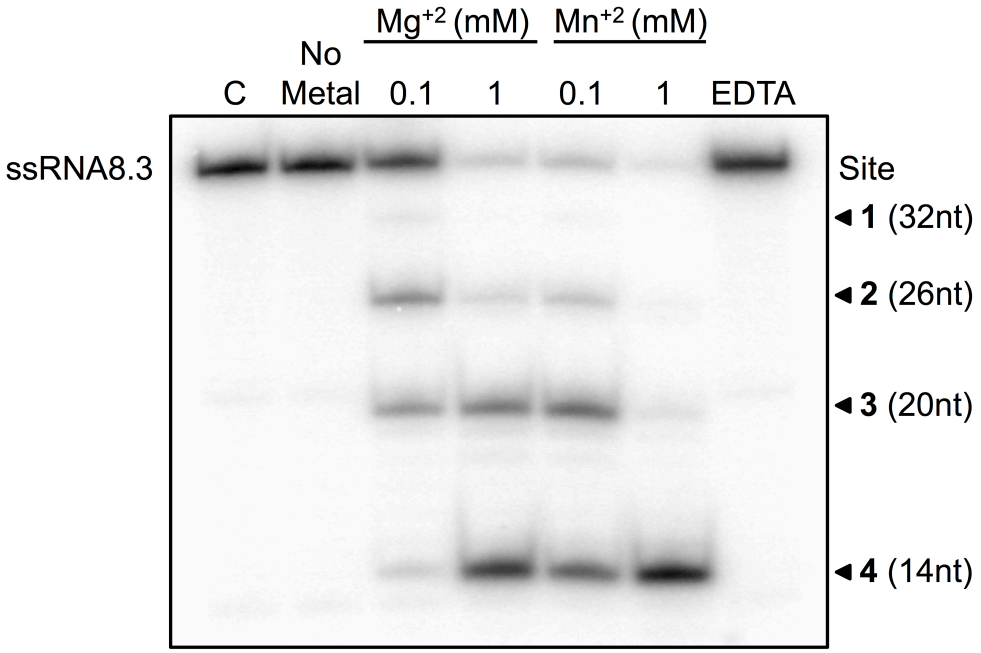


Figure 3.2 RNA cleavage by the Cmr complex is metal dependent. ssRNA8.3 target was incubated in the presence of the Cmr complex/crRNA8.3 with $MgCl_2$, $MnCl_2$, without metal, and with EDTA.

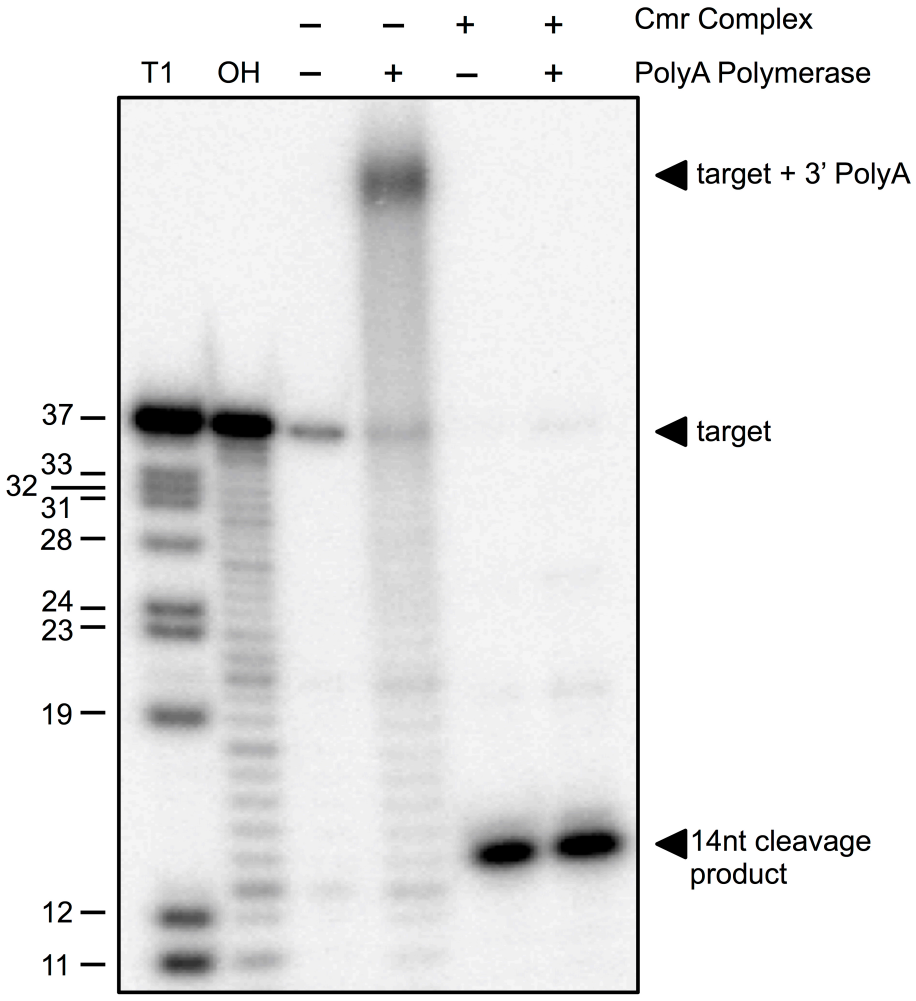


Figure 3.3 5' end-labeled target RNA and Cmr complex cleavage products were subject to polyadenylation by *E. coli* polyA polymerase (+) or no treatment (–). The position of the target RNA and polyadenylated target RNA are indicated. Also indicated is the location of the cleavage product. No polyadenylation of the cleavage products was observed, indicating the lack of 3' hydroxy groups on these RNAs.

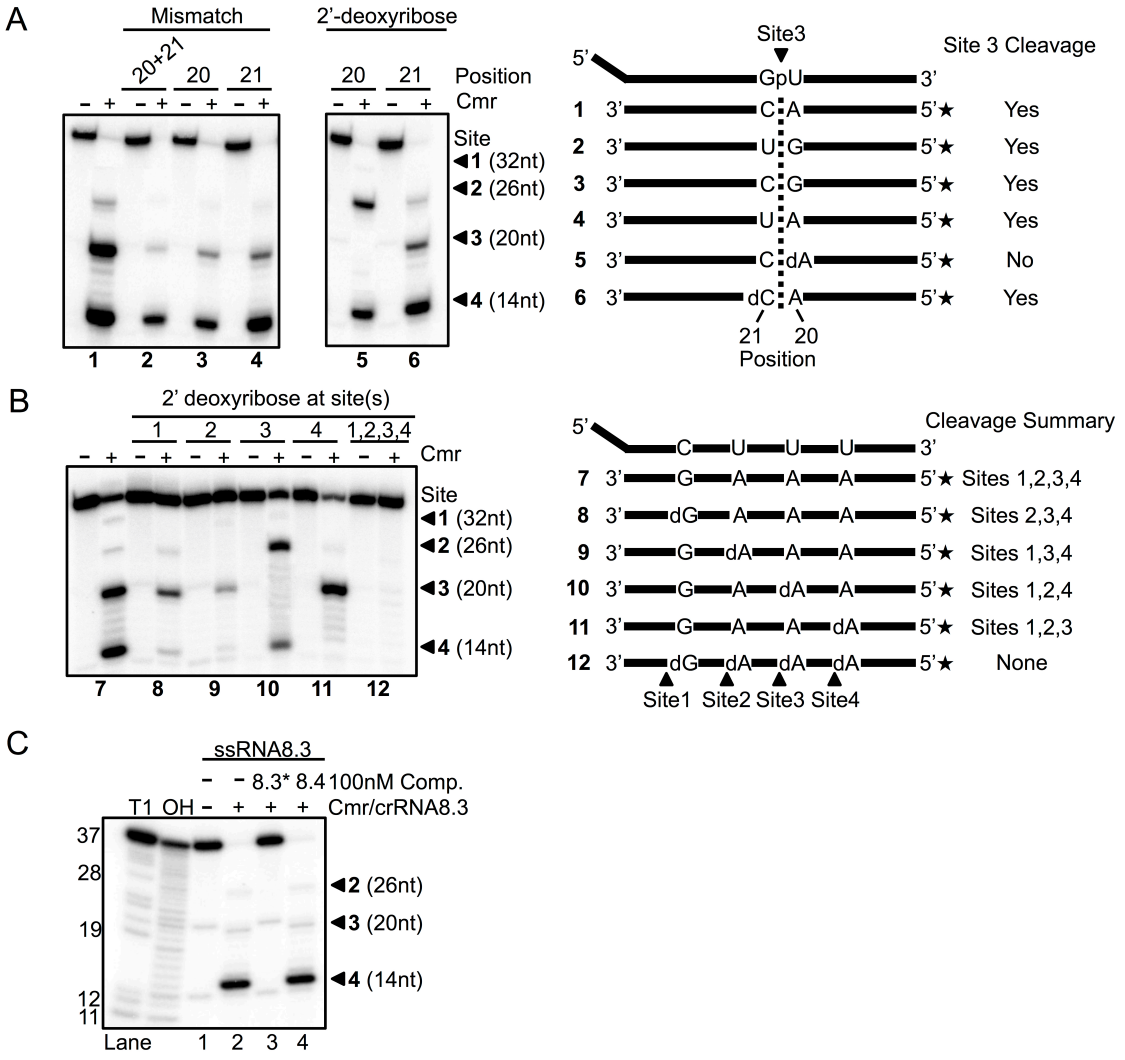


Figure 3.4 Determinants for RNA cleavage by the *Thermotoga maritima* Cmr Complex. Modified ssRNA8.3 targets were incubated with the Cmr complex and the products resolved on a denaturing polyacrylamide gel. (A) ssRNA targets were modified to contain either mismatches or 2'- deoxyribonucleotides at the indicated positions. A schematic of the modified targets is shown on the right. (B) Modified ssRNA targets with

2’-deoxyribonucleotide blocks at the indicated cleavage site(s). A schematic of the modified targets is shown on the right. (C) Completion experiments where the Cmr complex was incubated with the ssRNA8.3 target and either no competitor, a specific competitor (ssRNA8.3*) or an un-specific competitor (crRNA8.4)

Table 3.1 Sequence of RNA oligonucleotides used in assays

Name	Sequence (5'-3')
guide crRNA8.3	AUGGAAACCCGUCCCUUCUUUCCUGUCUUUUUGCCUCCUUUAAGU
guide crRNA8.3 (no psi-tag)	CCGUCCCUUCUUUCCUGUCUUUUUGCCUCCUUUAAGU
guide crRNA8.3 (scramble psi-tag)	GCAAUAGACCGUCCCUUCUUUCCUGUCUUUUUGCCUCCUUUAAGU
ssRNA8.3	ACUUAAAGGAGGCAAAAAGACAGGAAAGAAGGGACGG
dsRNA8.3	5'ACUUAAAGGAGGCAAAAAGACAGGAAAGAAGGGACGG3' 3’UGAAUUUCCUCCGUUUUUCUGUCCUUUCUUCCCUGCC5’
guide crRNA8.4	AUGGAAACUUGUACAAAGCUUCUACGUCAGGUUCAAUCCUGAUUU
ssRNA8.4	AAAUCAGGAUUGAACCUGACGUAGAAGCUUUGUACAA
Reverse ssRNA8.4	AACAUGUUUCGAAGAUGCAGUCCAAGUUAGGACUAAA
Mismatch 5'p3'	GGCAGGGAAGAAAGGAUGGAAAAACGGAGGAAAUUCA
Mismatch 5'	GGCAGGGAAGAAAGGACGGAAAAACGGAGGAAAUUCA
Mismatch 3'	GGCAGGGAAGAAAGGAUAGAAAAACGGAGGAAAUUCA

CHAPTER 4

DNA DEGRADATION BY THE CRISPR TYPE III-B COMPLEX

DNA cleavage by the Cmr complex

Genetic experiments implicate the Cmr complex in targeting transcriptionally active plasmid DNA in cells (Deng et al. 2013). Yet, Cmr complexes have only been observed to cleave RNA and not DNA *in vitro* (Figure 3.1A) (Hale et al. 2009; Staals et al. 2013; Zhang et al. 2012) and *in vivo* (Hale et al. 2012; Zebec et al. 2014).We reasoned that these differing observations could be reconciled if the DNA nuclease activity of the Cmr complex required the presence of a ssRNA transcript. To test this we incubated the *Thermotoga maritima* Cmr complex with crRNA8.3, an ssRNA8.3 target (a mimicking a transcript) and one of either of two different ssDNA substrates (substrate A and B), which were 5'-labeled (Figure 4.1A). The ssDNA substrates we selected were chosen at random and are not related at all in sequence to the crRNA or target RNA. We initially screened ssDNA as a potential substrate because Csm1 (which like Cmr2 is a Cas10 family protein) has ssDNA nuclease activity (Jung et al. 2015) and is the only subunit of a Type III effector complex with known DNA nuclease activity. The Cmr complex cut the DNA substrate weakly at multiple sites only when it and ssRNA8.3 target were added to the reaction at the same time (lanes 4 and 11 in Figure 4.1B). No cleavage was observed if the Cmr complex was pre-incubated with the ssRNA8.3 target for 10 minutes before addition of the ssDNA (lanes 3 and 10 in Figure 4.1B). This suggested cleavage of the ssRNA target may inhibit cleavage of the ssDNA substrate. No DNA cleavage was observed in the absence of the Cmr complex, in the absence of the complementary ssRNA8.3 target (Figure 4.1B) or with an ssRNA target non-complementary to

crRNA8.3 (lanes 7 and 14 in Figure 4.1B). Cleavage was dependent on the presence of manganese ions but not magnesium ions (Figure 4.2) and was specific to ssDNA as no cleavage of either ssRNA or dsDNA (Figure 4.1A) was observed under the same reactions conditions (lanes 15-18 in Figure 4.1B).

To further investigate the effects of RNA cleavage on the DNA cleavage activity of the Cmr complex, we monitored DNA cleavage under two regimes where the ssRNA target is not cut. Firstly, we monitored DNA cleavage using the non-cleavable ssRNA8.3* as the RNA target. Secondly, we monitored DNA cleavage using a Cmr complex that was reconstituted with Cmr4 harboring a D26A mutation (Benda et al. 2014; Ramia, Spilman, et al. 2014a; Zhu & Ye 2015). This variant of the *Thermotoga maritima* Cmr complex binds to (lane 3 in Figure 4.3) but is unable to cleave a complementary RNA target (lane 4 in Figure 4.4B). Under both of these regimes cleavage of the ssDNA was greatly enhanced (lanes 5-6 and lanes 12-13 in Figure 4.1B) suggesting that cleavage of the RNA target inhibits cleavage of ssDNA by the Cmr complex.

RNA target binding licenses ssDNA cleavage

Our results suggest that RNA target binding licenses ssDNA cleavage by the Cmr complex. To gain additional insight into this licensing we investigated RNA cleavage more closely. First, we wished to determine if the RNA products dissociate from the Cmr

complex following cleavage or if they remain bound. Thus, Cmr complex pre-equilibrated with an equimolar amount of crRNA8.3 was incubated with either equimolar, or 2- or 4-fold excess of the ssRNA8.3 target. Cleavage was monitored over time by denaturing PAGE followed by autoradiography. If RNA products do not dissociate from the complex following cleavage then in reactions with target in excess the amount of target cleaved should equal the amount of Cmr complex in the reaction. However, at all ratios of Cmr complex to target tested approximately 100% of the ssRNA8.3 target was cleaved (Figure 4.5A), indicating that the RNA products dissociate from the complex following cleavage. In a second set of experiments, the Cmr complex, pre-equilibrated with crRNA8.3 (1:1 stoichiometry) was incubated with 100 nM of target ssRNA8.3 and 1 nM of ssDNA substrate A (5'-radiolabeled). Cleavage of the DNA was monitored over time until the accumulation of product had plateaued (Figure 4.5B). Note that under these conditions (100nM of ssRNA8.3 target), the DNA cleavage activity of the Cmr complex is more robust than in previous experiments, where the concentration of the ssRNA8.3 target was much lower (1nM). Once the accumulation of product had plateaued either fresh complementary (ssRNA8.3) or in a control experiment non-complementary (ssRNA8.4) RNA targets were added to the reactions and we continued to monitor DNA cleavage (Figure 4.5B). Addition of fresh complementary ssRNA8.3 re-initiated cleavage of the ssDNA, whereas the non-complementary ssRNA8.4 had no effect, suggesting that the RNA products had dissociated from the complex. If the RNA products aren't released after cleavage then addition of fresh target should have no effect. Together these data indicate that the presence of an intact ssRNA target complementary

to the crRNA activates a ssDNA nuclease activity in the Cmr complex, and that cleavage of the ssRNA target and dissociation of the resulting fragments then prevent DNA cleavage.

In Type III-A systems base-pairing between the 5'-handle of the crRNA and the 3'-flanking sequence of the RNA target has been shown to have no effect on RNA target cleavage but inhibits cleavage of the DNA (Samai et al. 2015; Marraffini & Sontheimer 2010b). We tested the effects of complementarity to the 5'-handle on RNA and DNA cleavage by the *Thermotoga maritima* Cmr complex and found that the RNA was cleaved, albeit to a lesser extent (lane 5 in Figure 4.4A), and that cleavage of DNA was almost undetectable (lane 5 in Figure 4.4B). Thus, activation of DNA cleavage by the Cmr complex also requires a lack of complementarity between the RNA target and the crRNA 5'-handle.

Sequence specificity for the ssDNA target

The Cmr complex was able to cleave two ssDNA substrates with distinct sequences suggesting that the ssDNA nuclease activity of the Cmr complex is not sequence specific. However, cleavage of each substrate produced a distinct band pattern (Figure 4.1B) indicating some sequence preference at the sites of cleavage. To investigate this further we mapped the length of the cleavage products for the two different ssDNA substrates (Figure 4.1A). This mapping revealed the Cmr complex cleaved the DNA after every thymidine and not after any other nucleotide. Thymidine specificity was further

verified using a series of DNA substrates (derived from substrate A) where cleavage was blocked at the sites where thymidines were replaced with adenosine (Figure 4.6A). A ssDNA substrate completely lacking thymidine was not cleaved by the Cmr complex (Figure 4.6A).

In addition to the observed thymidine specificity, cleavage between a thymidine and a cytidine appeared to be less efficient than cleavage between a thymidine and any other nucleotide (see the 17-nt product of substrate A and the 22 nt product of substrate B in Figure 4.1B). Substrate A contains a single thymidine-cytidine dinucleotide sequence, located at positions 17-18 (Figure 4.2C). We mutated this cytidine (C18) to each of the other three nucleotides and monitored cleavage of the modified DNA substrate. Cleavage at position 17 of these modified substrates was more robust than in substrate A (Figure 4.6B). Thus, confirming that cleavage between a thymidine and a cytidine is less efficient than cleavage between thymidine and any other nucleotide.

Cleavage of ssDNA in the context of dsDNA

To determine if the *Thermotoga maritima* Cmr complex can cleave ssDNA in the context of dsDNA we generated a series of dsDNA substrates that contained mismatched "bubble" regions. The "bubble" regions ranged in size from 2-10 nt and consisted of a poly-thymidine sequence on one strand (the T-strand) and a poly-cytidine sequence on the opposite strand (the C-strand) (Figure 4.6C). The double-stranded regions of these

substrates were designed such that they did not melt at the reaction temperature, 80°C. The "bubble" substrates (labeled at the 5'-end of their T-strands) were incubated with the *Thermotoga maritima* Cmr complex, crRNA8.3 and an ssRNA8.3 target and then analyzed by denaturing PAGE and autoradiography. Each of the substrates was cleaved multiple times within the bubble region (Figure 4.6C) suggesting the Cmr complex can cleave ssDNA regions as short as 2-nt in the context of dsDNA.

Mutations in the HD domain of Cmr2 abolish ssDNA cleavage

The ssDNA nuclease activity of isolated Csm1 has been attributed to its HD domain (Jung et al. 2015).We reasoned that the HD domain of Cmr2 might be responsible for the observed ssDNA cleavage activity of the Cmr complex. To test this prediction, we made a mutant Cmr2 in which two highly conserved residues in the HD motif (His32 and Asp33) were substituted with alanine. The mutant Cmr2 (the HD-AA mutant) was then assembled into the Cmr complex and incubated with crRNA8.3 and either a 5'-labeled ssRNA8.3 target or a 5'-labeled ssDNA and an unlabeled ssRNA8.3 target. After a 10 minute incubation at 80°C, the HD-AA mutant complex did not cleave the ssDNA (lane 3 in Figure 4.7), but did cleave the ssRNA8.3 target to a 14 nt product (Figure 4.7). These observations suggest that the conserved HD-motif found in Cmr2 is critical for the ssDNA endonuclease activity of the Cmr complex.

In the Type III-A system of *Staphylococcus epidermidis* the GGDD motif in the palm domain of Csm1 is required for DNA cleavage and not the HD domain (Hatoum-Aslan et al. 2014; Samai et al. 2015; Ramia, Tang, et al. 2014b). To test if the GGDD motif has a role in DNA cleavage by the *Thermotoga maritima* Cmr complex we made a mutant Cmr2 in which the two highly conserved aspartate residues in the GGDD motif were substituted with alanine. The mutant Cmr2 (the GGDD-GGAA mutant) was then assembled into the Cmr complex and its cleavage activities were accessed as before with the HD-AA mutant. The GGDD-GGAA mutant complex failed to cleave ssDNA (lane 4 in Figure 4.7) but also failed to cleave the complementary ssRNA8.3 target (lane 8 in Figure 4.7). To further investigate why this mutant failed to cleave the RNA target, we used our electrophoretic mobility shift assay to access its specific binding to a complementary RNA target (lanes 5 and 10 in Figure 4.3). We found that the GGDD-GGAA mutant complex did not bind a complementary RNA target. This suggests that mutation of the GGDD motif in *Thermotoga maritima* Cmr2 subunit results in a dysfunctional Cmr complex that cannot bind complementary RNA targets and thus cannot be activated for DNA cleavage.

Materials and methods

Preparation of labeled oligonucleotides

DNA oligonucleotides were purchased from Sigma Aldrich (Table 4.1). They were 5' radiolabeled with T4 polynucleotide kinase (New England BioLabs) and γ-^{32}P ATP (Perkin Elmer) in 1x T4 polynucleotide kinase buffer at 37°C for 30 min. The substrates were resolved on a denaturing polyacrylamide gel, visualized by autoradiography, excised from the gel and placed in a 0.5mL solution of 0.3M sodium acetate overnight at 4°C followed by ethanol precipitation and resuspension in RNA Storage Solution (Ambion) for RNA. To generate dsDNA targets, labeled oligonucleotide with twice molar excess of complementary unlabeled oligonucleotide was incubated at 95°C for 10 min, followed by slow cooling to room temperature. Non-denaturing PAGE confirmed complete annealing.

DNA cleavage assay

Cmr2-6 was formed by pull-down (see Chapter 2 Materials and Methods). 250 nM Cmr2-6, 250 nM Cmr1 and 100 nM crRNA were incubated at 80°C for 30 min, in reaction buffer (100 mM KCl, 20 mM HEPES at pH 7, 1 mM $MnCl_2$, 0.5 mM ATP and 1 mM TCEP). This complex was then incubated with both a target RNA (1nM, unless

otherwise indicated) and 5'-labeled ssDNA (1nM, unless otherwise indicated) substrate in reaction buffer for 10 min at 80°C. The reaction was stopped at 10 min or the indicated time point by performing a phenol-chloroform extraction. The resulting water soluble fraction is added to equal parts quencher dye (90% formamide, 2.5% glycerol, 0.01% SDS, 0.01% bromophenol blue, 0.01% xylene cyanol and 1 mM EDTA) and heated at 95°C for 10 min. Radiolabeled DNA ladders were generated by purchasing oligonucleotides of defined sizes and 5' labeling them as above. The samples are then run on polyacrylamide denaturing gels and visualized by phosphorimaging.

Electrophoretic mobility shift assay

1nM of 5'-labeled RNA target was incubated with 800nM Cmr complex (wild-type and mutants complexes) loaded with 800nM crRNA8.3 for 1 hour at 4°C in reaction buffer. RNA cleavage is undetectable under these conditions. Following incubation bound RNA target was resolved from free target by electrophoresis through a 5% polyacrylamide gel. RNA was visualized by autoradiography.

Figures

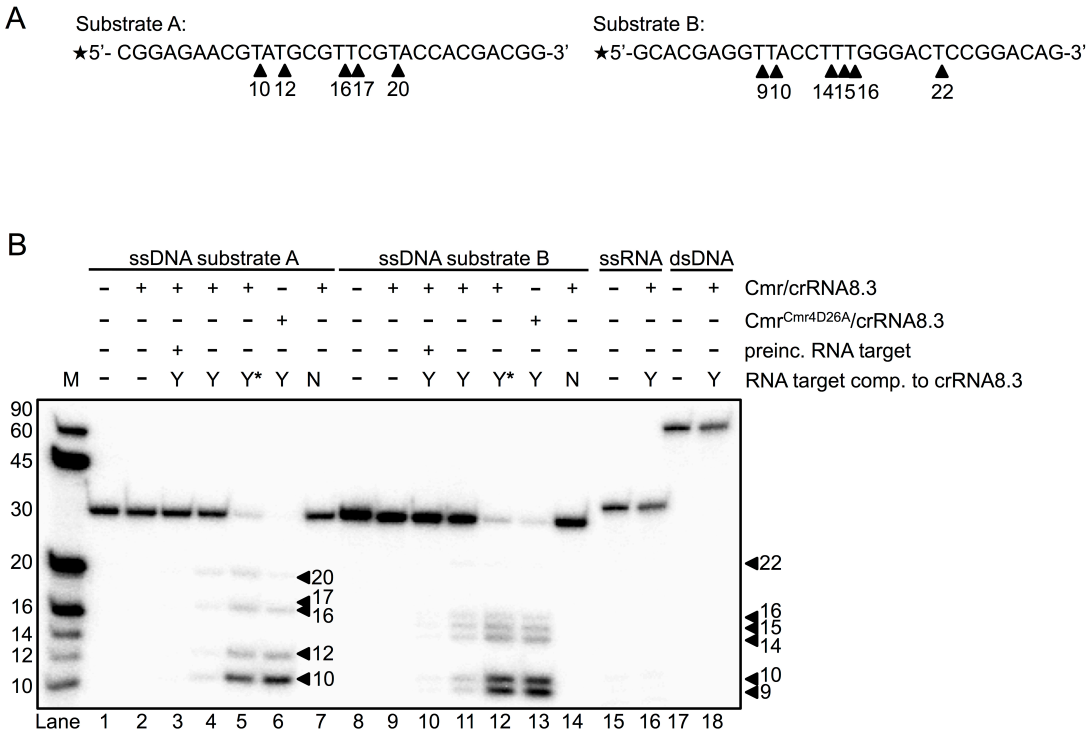


Figure 4.1 DNA cleavage by the Cmr complex. (A) 5'-labeled ssDNA substrates (substrates A and B), an ssRNA (whose sequence is equivalent to ssDNA substrate A) or a dsDNA (substrate A annealed to a complementary oligo) were incubated with the Cmr complex in the presence of ssRNA8.3 targets and then analyzed by denaturing PAGE. A "*" indicates the use of the non-cleavable ssRNA8.3* target, "8.4" indicates the use of the ssRNA8.4 target. M denotes the marker lane. (B) Schematic showing the mapping of the cleavage sites onto the sequences of ssDNA substrates A and B.

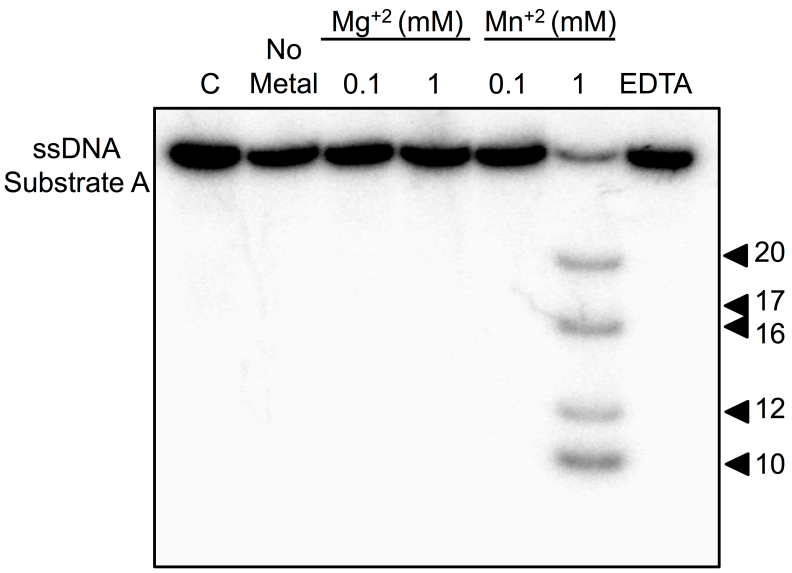


Figure 4.2 DNA cleavage by the Cmr complex is metal dependent. ssDNA Substrate A is incubated in the presence of the Cmr complex with $MgCl_2$, $MnCl_2$, without metal, and with EDTA.

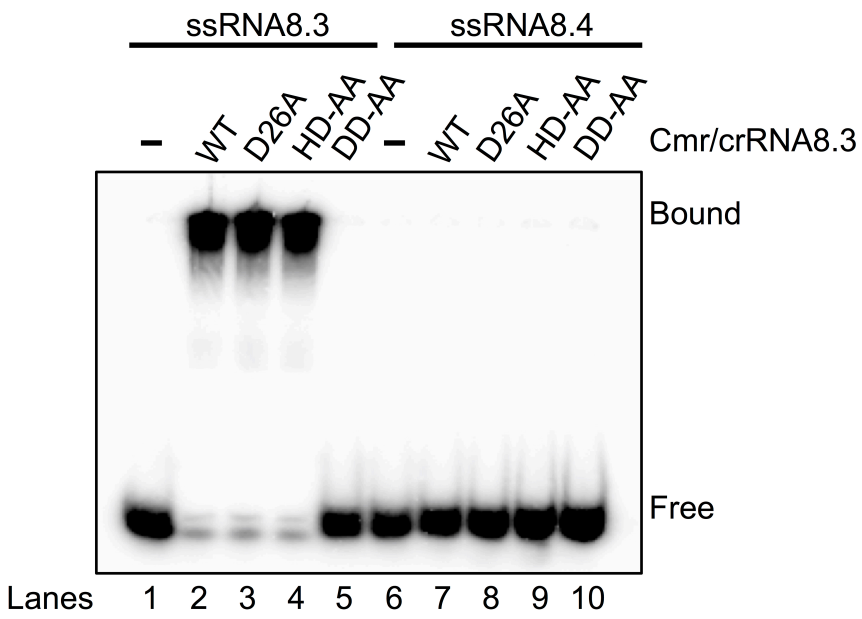


Figure 4.3 Target ssRNA binding by wild-type and mutant Cmr/crRNA8.3 complexes. The indicated recombinant Cmr complexes, with crRNA8.3 used as a guide, were incubated with 5' end-labeled 37-nt single-stranded target RNA that was either complementary (ssRNA8.3) or not (ssRNA8.4) to the crRNA. RNA and complexes were resolved by native gel electrophoresis. As indicated, the bound target RNAs will migrate slower through the gel.

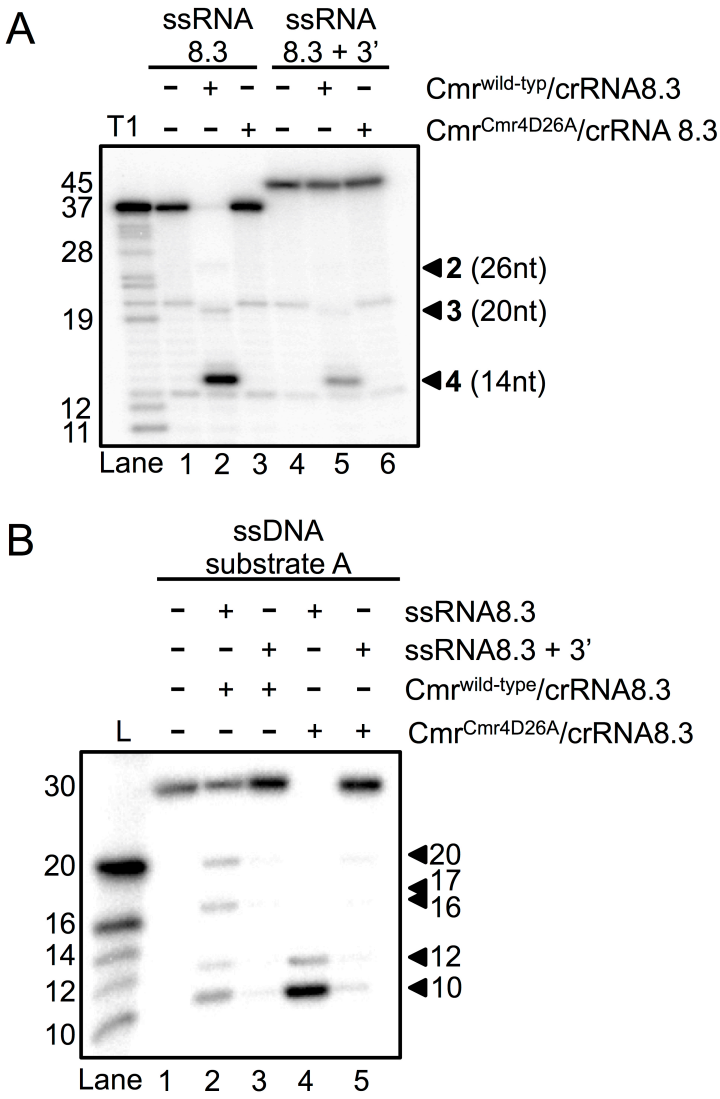


Figure 4.4 Mutation of the D26A residue in Cmr4 eliminates ssRNA cleavage and complementarity by the target RNA to the crRNA at the 5' handle diminishes ssDNA activity. (A) Cleavage of 5' radiolabeled ssRNA8.3 and ssRNA8.3+3' incubated with wild-type or Cmr4 D26A containing Cmr complexes. ssRNA8.3+3' contains 8 nt on its

3’-end that are complementary to the 5’ handle of the crRNA. (B) 5’ radiolabeled ssDNA substrate A is mixed with the indicated Cmr complex containing either ssRNA8.3 or ssRNA8.3+3’. Cleavage products are resolved on denaturing polyacrylamide gels.

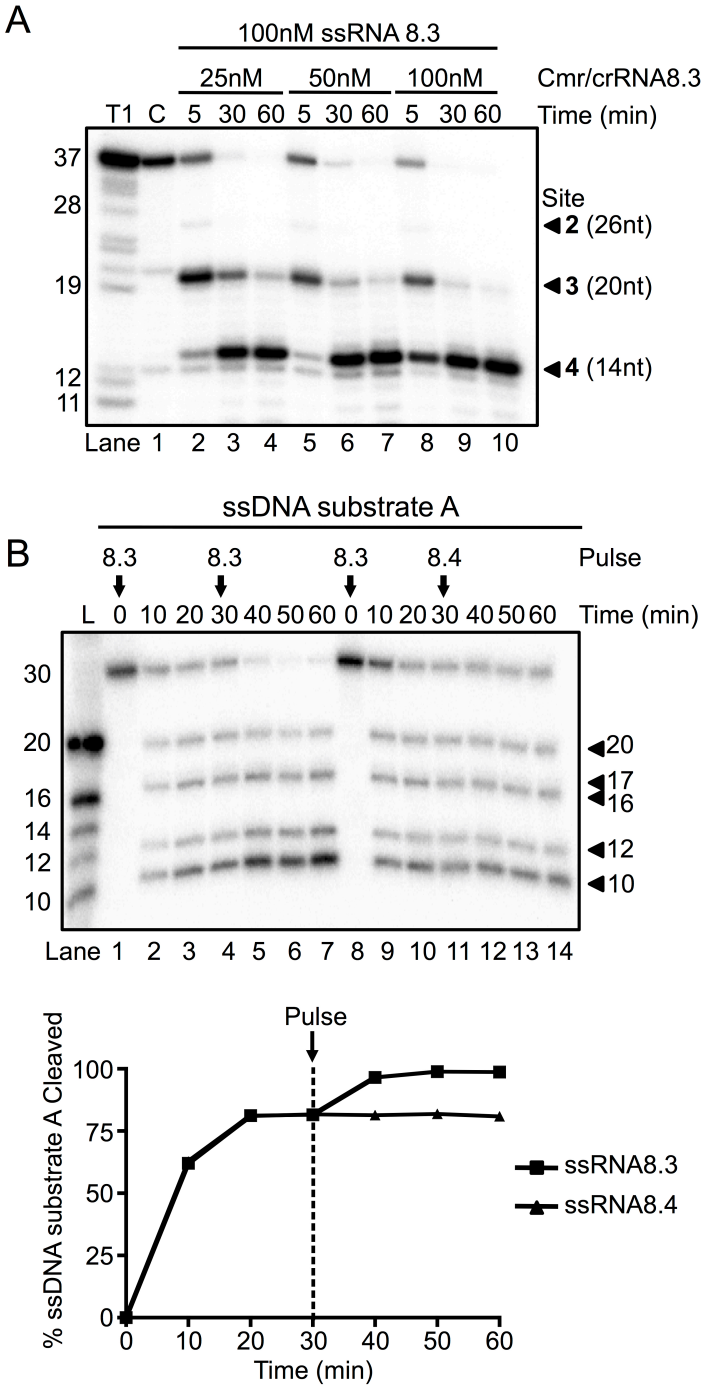


Figure 4.5 RNA turnover by the Cmr complex. (A) 5′ labeled ssRNA8.3 target was incubated with the Cmr complex in the presence of unlabeled crRNA8.3 at the indicated

concentrations, and cleavage was monitored over time. “T1” denotes the T1 lad- der, and “C” denotes a control lane containing only the labeled ssRNA8.3 target. (B) 5′ labeled ssDNA substrate A was incubated with the unlabeled ssRNA8.3, the Cmr complex, and crRNA8.3 for 30 min before either fresh ssRNA8.3 or ssRNA8.4 was added to the reaction (pulse), which was then monitored for a further 30 min. Samples were analyzed by denaturing PAGE. Below the gel panel is a quantification of the cleavage activity over time. The point at which the pulse was added is indicated.

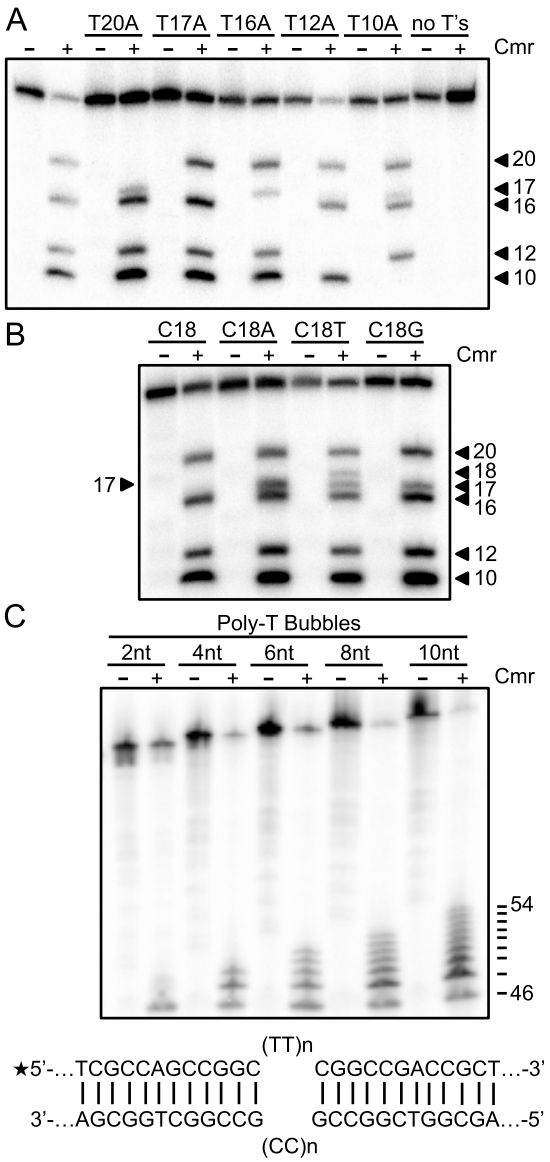


Figure 4.6 Determinants for DNA cleavage by the *Thermotoga maritima* Cmr complex. (A) 5'-labeled ssDNA substrates, with the indicated thymidine to adenosine substitutions, were incubated with (+) or without (-) the Cmr complex and analyzed by denaturing PAGE. (B) 5'-labeled ssDNA substrates, with the indicated substitutions at position 18, were incubated with (+) or without (-) the Cmr complex and analyzed by denaturing PAGE. (C) dsDNA substrates containing 2, 4, 6, 8 and 10 nt "bubbles" were incubated

with the Cmr complex and analyzed by denaturing PAGE. Below the gel is a schematic of the “bubble” substrate.

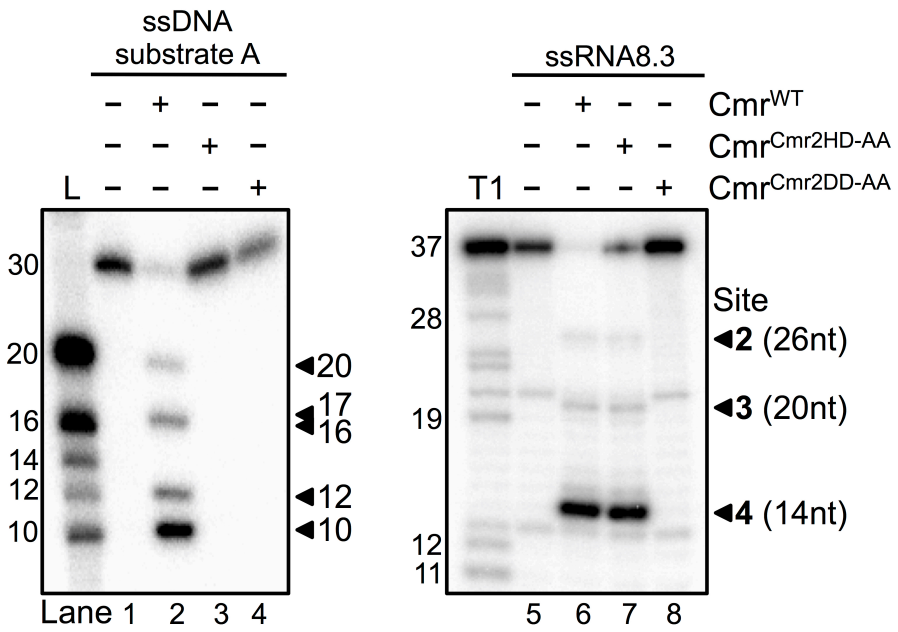


Figure 4.7 Effects of mutations in Cmr2. 5′ labeled ssRNA8.3 target (right) or 5′ labeled ssDNA substrate A with unlabeled ssRNA8.3 target (left) was incubated with either wild-type (WT) or mutant Cmr complexes (assembled with Cmr2 HD-AA or Cmr2 DD-AA) and then analyzed by denaturing PAGE. "L" denotes a DNA ladder, and "T1" denotes a T1 RNA ladder.

Table 4.1 Sequence of DNA oligonucleotides used in assays

Oligo Name	Sequence (5'-3')
ssDNA8.3	ACTTAAAGGAGGCAAAAAGACAGGAAAGAAGGGACGG
T20A	CGGAGAACGTATGCGTTCGAACCACGACGG
T17A	CGGAGAACGTATGCGTACGTACCACGACGG
T16A	CGGAGAACGTATGCGATCGTACCACGACGG
T12A	CGGAGAACGTAAGCGTTCGTACCACGACGG
T10A	CGGAGAACGAATGCGTTCGTACCACGACGG
no T's	CGGAGAACGAAAGCGAACGAACCACGACGG
C18	CGGAGAACGTATGCGTTCGTACCACGACGG
C18A	CGGAGAACGTATGCGTTAGTACCACGACGG
C18T	CGGAGAACGTATGCGTTTGTACCACGACGG
C18G	CGGAGAACGTATGCGTTGGTACCACGACGG
dsDNA target	5'GCGCGCGCCGCGGGGGCCCGCCGGCGGGCGAGCTCGCCAGCCGGCCGGCCGACCGCTCGAGCGGGCGGCCGCCCGGGGGCGCCGCGCGCG3' 3’CGCGCGCGGCGCCCCCGGGCGGCCGCCCGCTCGAGCGGTCGGCCGGCCGGCTGGCGAGCTCGCCCGCCGGCGGGCCCCCGCGGCGCGCGC5’
2nt Poly-T Bubble	5'GCGCGCGCCGCGGGGGCCCGCCGGCGGGCGAGCTCGCCAGCCGGCTTCGGCCGACCGCTCGAGCGGGCGGCCGCCCGGGGGCGCCGCGCGCG3' 3’CGCGCGCGGCGCCCCCGGGCGGCCGCCCGCTCGAGCGGTCGGCCGCCGCCGGCTGGCGAGCTCGCCCGCCGGCGGGCCCCCGCGGCGCGCGC5’

4nt Poly-T Bubble	5'GCGCGCGCCGCGGGGGCCCGCCGGCGGGCGAGCT CGCCAGCCGGCTTTTCGGCCGACCGCTCGAGCGGG CGGCCGCCCGGGGGCGCCGCGCGCG3' 3’CGCGCGCGGCGCCCCCGGGCGGCCGCCCGCTCGA GCGGTCGGCCGCCCCGCCGGCTGGCGAGCTCGCCC GCCGGCGGGCCCCCGCGGCGCGCGC5’
6nt Poly-T Bubble	5'GCGCGCGCCGCGGGGGCCCGCCGGCGGGCGAGCT CGCCAGCCGGCTTTTTTCGGCCGACCGCTCGAGCGG GCGGCCGCCCGGGGGCGCCGCGCGCG3' 3’CGCGCGCGGCGCCCCCGGGCGGCCGCCCGCTCGA GCGGTCGGCCGCCCCCCGCCGGCTGGCGAGCTCGC CCGCCGGCGGGCCCCCGCGGCGCGCGC5’
8nt Poly-T Bubble	5'GCGCGCGCCGCGGGGGCCCGCCGGCGGGCGAGCT CGCCAGCCGGCTTTTTTTTCGGCCGACCGCTCGAGC GGGCGGCCGCCCGGGGGCGCCGCGCGCG3' 3’CGCGCGCGGCGCCCCCGGGCGGCCGCCCGCTCGA GCGGTCGGCCGCCCCCCCCGCCGGCTGGCGAGCTC GCCCGCCGGCGGGCCCCCGCGGCGCGCGC5’
10nt Poly-T Bubble	5'GCGCGCGCCGCGGGGGCCCGCCGGCGGGCGAGCT CGCCAGCCGGCTTTTTTTTTTCGGCCGACCGCTCGA GCGGGCGGCCGCCCGGGGGCGCCGCGCGCG3' 3’CGCGCGCGGCGCCCCCGGGCGGCCGCCCGCTCGA GCGGTCGGCCGCCCCCCCCCGCCGGCTGGCGAGC TCGCCCGCCGGCGGGCCCCCGCGGCGCGCGC5’

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

The Cmr complex from the Type III-B CRISPR-Cas system is a crRNA programmed effector complex that cleaves complementary RNA *in vitro* (Hale et al. 2009; Zhang et al. 2012; Staals et al. 2013) and *in vivo* (Hale et al. 2012; Zebec et al. 2014). The Cmr complex also protects against plasmid transformation *in vivo* in a transcription dependent manner (Deng et al. 2013). All known CRISPR-Cas systems that protect against DNA invasion do so by degrading the invading DNA (Garneau et al. 2010; Westra et al. 2012; Samai et al. 2015), suggesting that DNA targeting by the Cmr complex may also be mediated by DNA cleavage. Here we show the Cmr complex cleaves ssDNA in the presence of a complementary RNA target (Figure 4.1) and discuss the implications of this finding.

We demonstrate that Cmr proteins from *Thermotoga maritima* assemble into a Cmr complex (Figure 2.3 and Figure 2.4) with the same subunit interactions as those observed for Cmr complexes from other species (Spilman et al. 2013; Staals et al. 2013; Taylor et al. 2015; Benda et al. 2014; Osawa et al. 2015). *Thermotoga maritima* Cmr complex specifically cleaves RNA targets that are complementary to crRNA. Cleavage occurs sequentially over four sites, separated by 6 nt intervals (Figure 3.1C-E), and is catalyzed by the Cmr4 subunit (lane 4 in Figure) (Benda et al. 2014; Ramia, Spilman, et al. 2014a; Zhu & Ye 2015). As observed previously with a chimeric Cmr complex (Osawa et al. 2015) we demonstrate that cleavage is dependent on the 2' hydroxyl adjacent to the scissile phosphate (Figure 3.4). We also find that RNA cleavage by the Cmr complex can tolerate mismatches in the region of a cut site (Figure 3.4A). A

tolerance for mismatches was also observed with Csm complexes (Staals et al. 2014; Tamulaitis et al. 2014), suggesting that both Type III-A and Type III-B systems have flexibility in target specificity.

The Type III CRISPR-Cas systems protect cells from invasion by DNA viruses and plasmids in a transcription-coupled manner (Deng et al. 2013; Goldberg et al. 2014; Samai et al. 2015). Here we show that a purified Cmr complex can cleave ssDNA after thymidine nucleotides but only in the presence of a complementary ssRNA target (Figure 4.1). The *S. solfataricus* Cmr complex has been reported to cleave RNA targets at uridine-adenosine dinucleotides (Zhang et al. 2012), suggesting that sequence-specific cleavage of their substrates maybe a more general feature of Cmr complexes. RNA-activated DNA cleavage by the Cmr complex provides a basis for the previously observed transcription-coupled DNA targeting. Binding but not cleavage of the ssRNA target was required to activate DNA cleavage (Figs. 3, 4, S5 and S6), consistent with immunity against plasmid transformation in the Type III-A system requiring DNA, but not RNA cleavage (Samai et al. 2015). Although the structural basis of DNA nuclease activation is unknown, it may involve the conformational rearrangements observed on complementary target binding by cryo-electron microscopy (Taylor et al. 2015). We mapped the active site for ssDNA cleavage to the HD domain of Cmr2 (Figure 6A). Transcription-coupled DNA cleavage may therefore target regions of ssDNA that are generated by transcription, likely at the displaced non-template strand (Figure 7) (Samai

et al. 2015; Jackson & Wiedenheft 2015). Accordingly, we observe cleavage of dsDNA that contains a mismatched "bubble" region as small as two nucleotides (Figure 4.6C).

Several subunits of the Type I and Type III CRISPR-Cas complexes share distant phylogenetic relationships (Makarova et al. 2011; Makarova et al. 2015) and accumulating structural data suggests that these complexes may have evolved from a common ancestor, reviewed in (Jackson & Wiedenheft 2015). Our studies indicate that the HD domain of Cmr2 is the domain responsible for DNA cleavage in Type III-B systems. In Type I systems DNA is also cleaved by an HD domain, in the helicase-nuclease Cas3, providing a further evolutionary link between the two systems. Indeed, the HD domains of Cmr2 and Cas3 both depend on transition metals and not magnesium for catalysis and both cleave ssDNA (Mulepati & Bailey 2011). *S. thermophilus* Cas3 has also been shown to preferentially cleave after thymidine nucleotides (Sinkunas et al. 2013).

The immunity against DNA invasion provided by the Type III CRISPR-Cas systems is dependent on COG1517 proteins (Makarova et al. 2012), Csm6 in Type III-A systems (Hatoum-Aslan et al. 2014) and Csx1 in Type III-B systems (Deng et al. 2013). Yet, none of these proteins stably associate with their respective effector complexes (Hale et al. 2009; Hatoum-Aslan et al. 2014) nor are they required for RNA cleavage by either complex (Figure 3.1) (Hale et al. 2009; Zhang et al. 2012; Staals et al. 2013; Staals et al. 2014; Tamulaitis et al. 2014; Samai et al. 2015). We also show cleavage of ssDNA by RNA-activated Cmr complex does not require the presence of Csx1 (Figure 4.1, 4.2

and 4.3). Likewise, transcription-coupled cleavage of DNA by the Csm complex does not require Csm6 (Samai et al. 2015). Thus, Csx1 and Csm6 do not appear to have a direct role in either RNA or DNA cleavage by the Type III effector complexes. Further biochemical studies are needed to understand the function of these proteins, although interestingly sequence analysis suggests this family of proteins may be nucleases (Makarova et al. 2012).

Two models have been proposed for how a Type III effector complex could identify its DNA target using a crRNA. In the first model, crRNA pairs with the non-template strand that has been exposed by a transcription bubble (Deng et al. 2013; Goldberg et al. 2014; Samai et al. 2015). In the second, crRNA pairs with the nascent transcript (Goldberg et al. 2014; Jackson & Wiedenheft 2015). We observe that pairing of a crRNA with an RNA target activates ssDNA cleavage by the Cmr complex and the sequence specificity of this nuclease activity has no dependence on the crRNA sequence. Also, the *P. furiosus* Cmr complex does not bind ssDNA that is complementary to the crRNA but does bind complementary ssRNA (Hale et al. 2014). Similarly, *S. thermophilus* Csm complex binds complementary ssRNA 100 times tighter than ssDNA (Tamulaitis et al. 2014). Together these data support the second model in which the Type III effector complex (Cmr or Csm) binds to the nascent transcript (Figure 5.1). Once recruited by transcript binding, the ssDNA cleavage activity of the effector complex is activated. A likely substrate for this activity being the displaced ssDNA of the displaced non-template strand (Samai et al. 2015) of the downstream transcription bubble (Figure

7), which has been show by previous foot-printing assays to be more accessible than the template-strand in a transcription bubble (Wang & Landick 1997).

The active site for DNA cleavage may vary both between and within Type III-A and Type III-B systems. Through mutagenesis we mapped the DNA nuclease activity of the Type III-B Cmr complex to the HD motif of Cmr2 (Figure 4.3), as initially predicted from sequence analysis (Makarova et al. 2011). The HD domain of the Csm1 protein from the Type III-A system of *Thermococcus onnurineus* was also found to have DNA nuclease activity (Jung et al. 2015). This provides a link between the Type III and Type I systems as Type I systems also use HD domains, located in Cas3, to destroy their DNA targets (Westra et al. 2012; Mulepati & Bailey 2013; Sinkunas et al. 2013). However, in the Type III-A system of *Staphylococcus epidermidis* the HD motif of Csm1 is not required for DNA cleavage (Hatoum-Aslan et al. 2014). Instead, the GGDD motif in the palm domain of Csm1 is required (Hatoum-Aslan et al. 2014; Samai et al. 2015; Ramia, Tang, et al. 2014b). Moreover, some Type III-B systems contain Cmr2 proteins lacking HD domains, such as *T. thermophilus* (Staals et al. 2013). In these organisms the Type III-B system may either only target RNA or the GGDD motif may be used for DNA cleavage. More structural and biochemical studies are required to clarify the role of these two motifs (HD versus GGDD) in the different Type III systems.

Nucleases need to be tightly regulated so they cleave only the intended substrates, as any non-specific activity could be deleterious to the cell. The transcript-binding model

(Figure 5.1) provides a mechanism for both spatial and temporal regulation of DNA cleavage by the Cmr complex. Spatial regulation can be achieved by tethering the activated Cmr complex to a specific region of DNA that is actively transcribing, through base pairing between the crRNA and transcript. This spatial restraint would ensure that activated Cmr complex does not cleave at other transcription bubbles, such as those found at transcribing host genes. Thus, spatial regulation may provide a mechanism for immunity, specific cleavage of transcribing foreign DNA, rather than cell death, cleavage of transcribing host DNA. Targeting antisense transcription of host CRISPR loci, which has been observed in several species (Garrett et al. 2015; Hale et al. 2012), appears to be prevented by base paring between the 5'-handle of the crRNA and the complementary region of the resulting antisense transcript (Figure 4.4). Temporal regulation can be achieved by coupling DNA cleavage to the presence of an intact transcript, as once the transcript is cleaved the DNA nuclease activity is deactivated (Figure 4.1B). Due to this coupling of DNA and RNA cleavage activities, in cells the rate of transcript cleavage should to be slower than the rate of DNA cleavage in order for efficient DNA cleavage to occur. This isn't the case in our *in vitro* experiments as the rate of ssDNA cleavage is slower than the rate of ssRNA cleavage, compare lanes 8-10 in Figure 4.5A with lanes 1-4 in Figure 4.5B. However, in our *in vitro* assay we use ssDNA oligonucleotides as substrates. In cells, we propose that the relevant substrate is the non-template strand of the transcription bubble. This substrate may be a better substrate for the Cmr complex than ssDNA oligonucleotides and thus may be cleaved at a higher rate. Also, as discussed above, by targeting a transcribing region of DNA the Cmr complex is tethered to its DNA

substrate by its interactions with the nascent transcript. This tethering would increase the effective concentration of the DNA, again likely stimulating the rate of its cleavage. In addition to the possible effects on DNA cleavage, the rate of RNA cleavage in the context of a transcript emerging from a transcription bubble maybe slower than with a free RNA oligonucleotide substrate. Further experiments carefully examining these rates in the context of transcription-coupled targeting both *in vitro* and *in vivo* will be required to provide a complete mechanistic description of the system.

In summary, our studies provide biochemical evidence that the Type III-B system cleave ssDNA in the presence of an RNA target complementary to the crRNA. These results are consistent with previous results showing the Type III-B system protects cells from plasmid transformation in a transcription-coupled manner (Deng et al. 2013). Our data further resolves the ambiguity in activities reported for Type III systems and supports a unified mechanism for all Type III systems in which DNA nuclease activity is coupled to transcription (Samai et al. 2015).

Figures

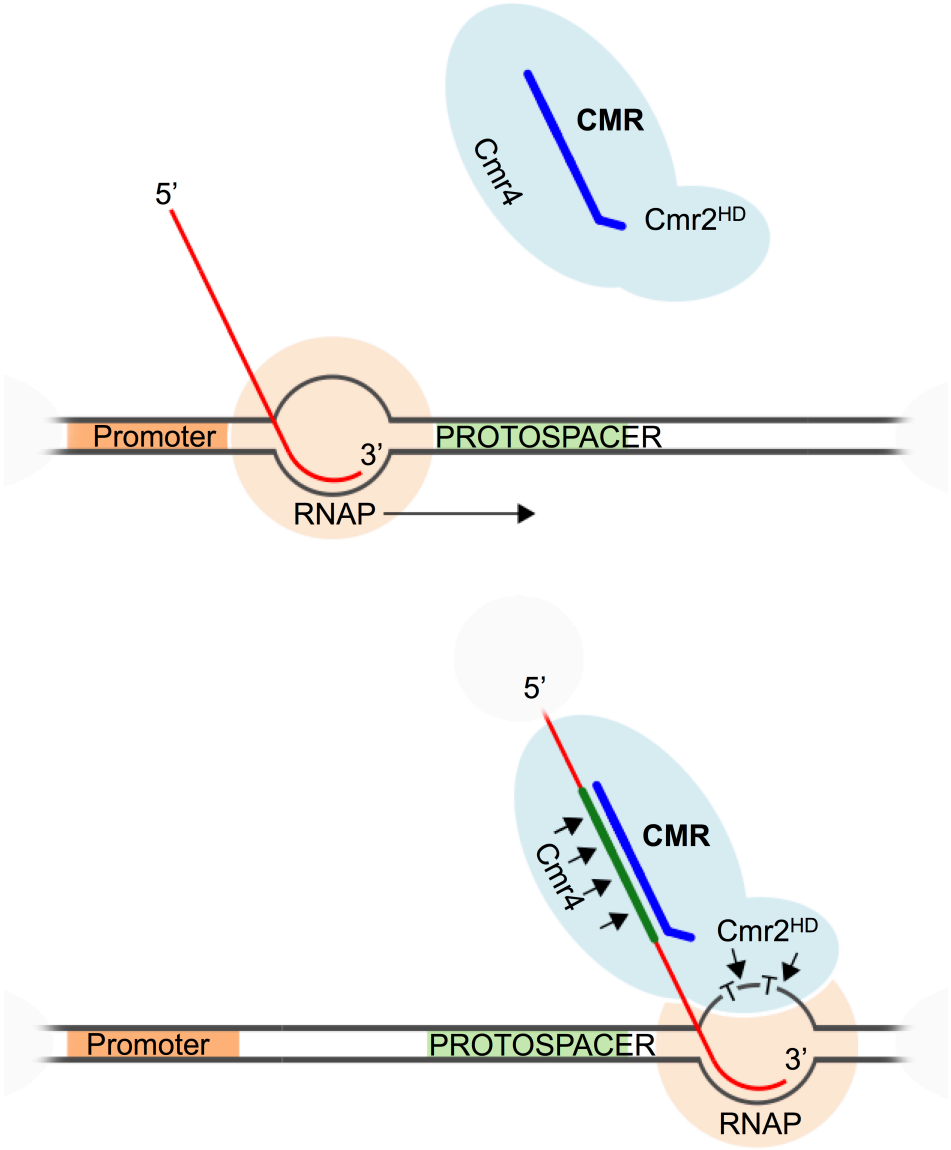


Figure 5.1 A model for the mode of action of the Type III-B CRISPR-Cas systems. Once RNA polymerase transcribes through the protospacer, the Cmr complex binds to the nascent transcript through base pairing with the crRNA. Transcript binding activates the

DNA and RNA nuclease activities of the Cmr complex. The DNA nuclease activity of the HD domain from Cmr2 cleaves the displaced non-template strand of the transcription bubble at sites 3' of thymidine nucleotides. The RNA nuclease activity cleaves the mRNA target stepwise through four sites. Activation of the DNA nuclease activity is dependent on the presence of a bound complementary target. Therefore, once the RNA target is cleaved the DNA nuclease activity is deactivated.

Bibliography

Barrangou, R. et al., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, 315(5819), pp.1709–1712.

Benda, C. et al., 2014. Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4. *Molecular cell*, 56(1), pp.43–54.

Bickle, T.A. & Krüger, D.H., 1993. Biology of DNA restriction. *Microbiological reviews*, 57(2), pp.434–450.

Bolotin, A. et al., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)*, 151(Pt 8), pp.2551–2561.

Breitbart, M. & Rohwer, F., 2005. Here a virus, there a virus, everywhere the same virus? *Trends in microbiology*, 13(6), pp.278–284.

Brouns, S.J.J. et al., 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (New York, N.Y.)*, 321(5891), pp.960–964.

Carte, J. et al., 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & development*, 22(24), pp.3489–3496.

Carte, J. et al., 2014. The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in Streptococcus thermophilus. *Molecular microbiology*, 93(1), pp.98–112.

Chopin, M.-C., Chopin, A. & Bidnenko, E., 2005. Phage abortive infection in lactococci: variations on a theme. *Current opinion in microbiology*, 8(4), pp.473–479.

Cocozaki, A.I. et al., 2012. Structure of the Cmr2 subunit of the CRISPR-Cas RNA silencing complex. *Structure (London, England : 1993)*, 20(3), pp.545–553.

Curtis, T.P., Sloan, W.T. & Scannell, J.W., 2002. Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16), pp.10494–10499.

Deltcheva, E. et al., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), pp.602–607.

Deng, L. et al., 2013. A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. *Molecular microbiology*, 87(5), pp.1088–1099.

Doolittle, W.F. & Zhaxybayeva, O., 2009. On the origin of prokaryotic species. *Genome research*, 19(5), pp.744–756.

Forde, A. & Fitzgerald, G.F., 1999. Bacteriophage defence systems in lactic acid bacteria. *Antonie van Leeuwenhoek*, 76(1-4), pp.89–113.

Garneau, J.E. et al., 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320), pp.67–71.

Garrett, R.A. et al., 2015. CRISPR-Cas Adaptive Immune Systems of the Sulfolobales: Unravelling Their Complexity and Diversity. *Life (Basel, Switzerland)*, 5(1), pp.783–817.

Gasiunas, G. et al., 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 109(39), pp.E2579–86.

Goldberg, G.W. et al., 2014. Conditional tolerance of temperate phages via transcription-

dependent CRISPR-Cas targeting. *Nature*, 514(7524), pp.633–637.

Grissa, I., Vergnaud, G. & Pourcel, C., 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC bioinformatics*, 8(1), p.172.

Haft, D.H. et al., 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology*, 1(6), p.e60.

Hale, C.R. et al., 2012. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Molecular cell*, 45(3), pp.292–302.

Hale, C.R. et al., 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, 139(5), pp.945–956.

Hale, C.R. et al., 2014. Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex. *Genes & development*, 28(21), pp.2432–2443.

Hatoum-Aslan, A. et al., 2014. Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *Journal of bacteriology*, 196(2), pp.310–317.

Huang, Z. & Szostak, J.W., 1996. A simple method for 3'-labeling of RNA. *Nucleic acids research*, 24(21), pp.4360–4361.

Huber R, Langworthy TA, König H, Thomm M, Woese CR, Sleytr UB, Stetter KO. 1986. Thermotoga maritima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch Microbiol* **144**: 324–333.

Ishino, Y. et al., 1987. Nucleotide sequence of the iap gene, responsible for alkaline

phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*, 169(12), pp.5429–5433.

Jackson, R.N. & Wiedenheft, B., 2015. A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses. *Molecular cell*, 58(5), pp.722–728.

Jansen, R. et al., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, 43(6), pp.1565–1575.

Jinek, M. et al., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096), pp.816–821.

Jore, M.M. et al., 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology*, 18(5), pp.529–536.

Jung, T.-Y. et al., 2015. Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity. *Structure (London, England : 1993)*, 23(4), pp.782–790.

Majumdar, S. et al., 2015. Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus. *RNA (New York, N.Y.)*, 21(6), pp.1147–1158.

Makarova, K.S. et al., 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct*, 1(1), p.7.

Makarova, K.S. et al., 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology*, 13(11), pp.722–736.

Makarova, K.S. et al., 2011. Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology*, 9(6), pp.467–477.

Makarova, K.S. et al., 2012. Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. *Biology direct*, 7(1), p.40.

Marraffini, L.A. & Sontheimer, E.J., 2010a. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature reviews. Genetics*, 11(3), pp.181–190.

Marraffini, L.A. & Sontheimer, E.J., 2010b. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, 463(7280), pp.568–571.

Mojica, F.J. et al., 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology*, 36(1), pp.244–246.

Mojica, F.J.M. et al., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, 60(2), pp.174–182.

Mulepati, S. & Bailey, S., 2013. In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *The Journal of biological chemistry*, 288(31), pp.22184–22192.

Mulepati, S. & Bailey, S., 2011. Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *The Journal of biological chemistry*, 286(36), pp.31896–31903.

Nelson, K.E. et al., 1999. Evidence for lateral gene transfer between Archaea and bacteria

from genome sequence of Thermotoga maritima. *Nature*, 399(6734), pp.323–329.

Osawa, T. et al., 2015. Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. *Molecular cell*, 58(3), pp.418–430.

Peränen J, Rikkonen M, Hyvönen M, Kääriäinen L. 1996. T7 Vectors with a Modified T7lacPromoter for Expression of Proteins inEscherichia coli. *Analytical Biochemistry* **236**: 371–373.

Pourcel, C., Salvignol, G. & Vergnaud, G., 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading, England)*, 151(Pt 3), pp.653–663.

Ramia, N.F., Spilman, M., et al., 2014a. Essential structural and functional roles of the Cmr4 subunit in RNA cleavage by the Cmr CRISPR-Cas complex. *Cell reports*, 9(5), pp.1610–1617.

Ramia, N.F., Tang, L., et al., 2014b. Staphylococcus epidermidis Csm1 is a 3"-5" exonuclease. *Nucleic acids research*, 42(2), pp.1129–1138.

Rath, D. et al., 2015. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie*, 117, pp.119–128.

Rouillon, C. et al., 2013. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Molecular cell*, 52(1), pp.124–134. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=24119402&retmode=ref&cmd=prlinks.

Samai, P. et al., 2015. Co-transcriptional DNA and RNA Cleavage during Type III

CRISPR-Cas Immunity. *Cell*, 161(5), pp.1164–1174.

Seed, K.D., 2015. Battling Phages: How Bacteria Defend against Viral Attack. V. L. Miller, ed. *PLoS pathogens*, 11(6), p.e1004847.

Semenova, E. et al., 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America*, 108(25), pp.10098–10103.

Sinkunas, T. et al., 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *The EMBO journal*, 30(7), pp.1335–1342.

Sinkunas, T. et al., 2013. In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *The EMBO journal*, 32(3), pp.385–394.

Soucy, S.M., Huang, J. & Gogarten, J.P., 2015. Horizontal gene transfer: building the web of life. *Nature reviews. Genetics*, 16(8), pp.472–482.

Spilman, M. et al., 2013. Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Molecular cell*, 52(1), pp.146–152.

Staals, R.H.J. et al., 2014. RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. *Molecular cell*, 56(4), pp.518–530.

Staals, R.H.J. et al., 2013. Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. *Molecular cell*, 52(1), pp.135–145.

Stern, A. et al., 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends in genetics : TIG*, 26(8), pp.335–340.

Suttle, C.A., 2005. Viruses in the sea. *Nature*, 437(7057), pp.356–361.

Tamulaitis, G. et al., 2014. Programmable RNA shredding by the type III-A CRISPR-Cas system of Streptococcus thermophilus. *Molecular cell*, 56(4), pp.506–517.

Taylor, D.W. et al., 2015. Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science (New York, N.Y.)*, 348(6234), pp.581–585.

Thomas, C.M. & Nielsen, K.M., 2005. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature reviews. Microbiology*, 3(9), pp.711–721.

Tock, M.R. & Dryden, D.T.F., 2005. The biology of restriction and anti-restriction. *Current opinion in microbiology*, 8(4), pp.466–472.

van der Oost, J. et al., 2014. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature reviews. Microbiology*, 12(7), pp.479–492.

Wang, D. & Landick, R., 1997. Nuclease cleavage of the upstream half of the nontemplate strand DNA in an Escherichia coli transcription elongation complex causes upstream translocation and transcriptional arrest. *The Journal of biological chemistry*, 272(9), pp.5989–5994.

Westra, E.R. et al., 2012. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular cell*, 46(5), pp.595–605.

Whitman, W.B., Coleman, D.C. & Wiebe, W.J., 1998. Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, 95(12), pp.6578–6583.

Wommack, K.E. & Colwell, R.R., 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiology and molecular biology reviews : MMBR*, 64(1), pp.69–114.

Zebec, Z. et al., 2014. CRISPR-mediated targeted mRNA degradation in the archaeon Sulfolobus solfataricus. *Nucleic acids research*, 42(8), pp.5280–5288.

Zhang, J. et al., 2012. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Molecular cell*, 45(3), pp.303–313.

Zhu, X. & Ye, K., 2015. Cmr4 is the slicer in the RNA-targeting Cmr CRISPR complex. *Nucleic acids research*, 43(2), pp.1257–1267.

Curriculum Vitae

Michael A. Estrella

• 1731 Starr Street • Ridgewood, NY 11385 • michael.estrella@gmail.com •

Education

May 2016 — **The Johns Hopkins University** — Baltimore, MD
Doctor of Philosophy, GPA 3.61/4.0
Advisor: Scott Bailey, PhD
Thesis: *Nucleic Acid Degradation by the CRISPR Type III-B Complex*

May 2008 — **St. Francis College** — Brooklyn, NY
Bachelors of Science in Chemistry and Biology
Graduated Summa Cum Laude, GPA 3.87/4.0
Advisor: Filomena Califano, PhD

Research Experience

June 2010 – Jan 2016 — **Johns Hopkins Bloomberg School of Public Health** — Baltimore, MD
PhD Candidate, PI: Dr. Scott Bailey

- Used biochemical and biophysical techniques to study the CRISPR Type III-B targeting complex from *Thermotoga maritima*.
- Performed molecular cloning using traditional, ligation independent, and Gibson methods.
- Recombinant protein purification was performed on AKTA FPLC systems, using affinity, ion exchange, and size exclusion chromatography.
- Developed assays to characterize the Type III-B complex, including: pull-down, thermal melt, electrophoretic mobility shift, nuclease and foot printing assays.
- Determined the structure of two subunits from the Type III-B complex using protein crystallography, and experimental phasing was performed using selenomethionine and heavy metal soaks.
- Trained five PhD rotation students, a student volunteer, and a master's degree student.

June 2008 – Dec 2009 — **Brookhaven National Laboratory, Chemistry Department** — Upton, NY
Research Associate in Catalysis Group, PI: Dr. José A. Rodriguez

- Used synchrotron based techniques, X-ray diffraction (XRD) and X-ray absorption spectroscopy (XAS), to characterize structural properties of catalysts under *in situ* reaction conditions.
- Conducted gas chromatography and mass spectrometry analysis to study the catalytic performance for hydrogen production during Water Gas Shift (WGS) and EtOH steam reforming reactions.
- Prepared inorganic powder catalyst using Wet Impregnation and Deposition Precipitation.
- Trained and assisted users of X7B of the National Synchrotron Light Source (NSLS) as Beamline Technical Contact.
- Designed and built gas handling system for X7B allowing *in situ* catalytic test at high pressure and temperature.

- Involved in beamline development at X18A of the NSLS to test combined XRD/XAS and Diffuse Anomalous Fine Structure (DAFS) experiments.
- Conducted surface science experiments at U12A of the NSLS. X-ray photoelectron spectroscopy (XPS) and near edge X-ray absorption spectroscopy (NEXAFS) experiments were carried out at ultra-high vacuum (UHV) conditions on model metal surfaces.

June 2007 – Aug 2007 **Rohm & Haas, Paint and Coating Materials** Spring House, PA
Intern in Floor Care Division, Team Leader: Dr. Ted Tysak

- Assisted in upgrading an eco-friendly floor finish using Duragreen Polymer™.
- Investigated the synthesis and application of an experimental polymer to carpet care.

June 2006 – Dec 2006 **Princeton University, Dept. of Chemical Engineering** Princeton, NJ
Intern in Chemical Engineering Laboratory, PI: Dr. David W. Wood

- Investigated the non-chromatographic purification of recombinant proteins expressed in *E. coli* using two techniques; centrifugation and tangential flow filtration.
- Purification was based on a novel reversibly precipitating, self-cleaving tag.

Summer 2004 – 2005 **Wyckoff Heights Medical Center, Pathology Laboratory** Brooklyn, NY
Summer Intern, Supervisor: Dr. Vincenzo Napolitano

- Used hematology analyzer to conduct complete blood counts (CBC).
- Prepared tissue slides and blood smears and analyzed them using microscopy to observe morphology and histology.
- Conducted blood specific gravity testing, urine testing and pregnancy testing.

Teaching Experience

Spring 2012 – 2015 **Johns Hopkins Bloomberg School of Public Health** Baltimore, MD
Teaching Asst. for Genome Integrity and Cancer organized by Paul Miller

- Assisted in teaching graduate students molecular mechanisms devoted to the preservation of genome integrity in prokaryotic and eukaryotic cells.
- Organized one-on-one, small group, and whole-class reviews.
- Administered and graded exams.

Sept 2007 – May 2008 **The Research Foundation of the City University of NY** Jamaica, NY
Tutor in Chemistry and Lab Instructor

- Tutored high school seniors from the *Queens Bridge to Medicine Program* for acceptance into the Sophie Davis Medical School.
- Conducted collegiate chemistry labs for minority high school students to help them develop experimental techniques and aid them in interpreting and presenting experimental data.

Sept 2006 – May 2008 **St. Francis College, Academic Enhancement** Brooklyn, NY
Tutor in Chemistry and Biology

- Conducted lectures and one-on-one sessions, to teach underclassmen concepts in General Chemistry, Organic Chemistry, Biochemistry and General Biology.

Honors and Awards

2013 – 2016	Ruth L. Kirschstein NRSA Predoctoral Fellowship Award (NIGMS, F31GM105364)
2012 – 2016	Recipient of the Johns Hopkins Sommer Scholar Award
2008	Finalist in the Department of Energy SERCh poster competition
2008	Graduated Summa Cum Laude from St. Francis College
2008	Member of Chi Beta Phi National Science Honor Society
2008	Member of Duns Scotus Honor Society at St. Francis College
2007	Winner of NOBCChE Undergraduate Research Award
2006 – 2008	Recipient of Dean's Scholarship at St. Francis College
2006 – 2008	Recipient of the National SMART grant

Publications

1. **M. A. Estrella**, F. T. Kuo, and S. Bailey. *RNA Activated DNA cleavage by the Type III-B CRISPR-Cas effector complex.* Genes and Development, 2016. **30**: p. 460-470
2. L. Barrio, G. Zhou, I. D. González, **M. Estrella**, J. Hanson, J. A. Rodriguez, R. M. Navarro and J. L. G. Fierro. *In situ characterization of Pt catalysts supported on ceria modified TiO_2 for the WGS reaction: influence of ceria loading.* Physcial Chemistry Chemical Physics, 2012. **14**(7): p. 2192-2202.
3. A. I. Frenkel, Q. Wang, N. Marinkovic, J. G. Chen, L. Barrio, R. Si, A. López-Cámara, **M. A. Estrella**, J. A. Rodriguez and J. C. Hanson. *Combining X-Ray Absorption and X-Ray Diffraction Techniques for In Situ Studies of Chemical Transformations in Heterogeneous Catalysis: Advantages and Limitations.* Journal of Physical Chemistry C, 2011. **115**(36): p. 17884-17890.
4. L. Barrio, **M. Estrella**, J. C. Hanson and J. A. Rodriguez. *In-situ XRD study of the reduction of copper spinel with H2 and CO.* Acta Crystallographica Section A, 2011. **67**(A1): C139-C139
5. S.N. Ehrlich, J.C. Hanson, A. Lopez Camara, L. Barrio**, M. Estrella**, G. Zhou, R. Si, S. Khalid and Q. Wang. *Combined XRD and XAS*. Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment, 2011. **649**(1): p. 213-215.
6. G. Zhou, L. Barrio, S. Agnoli, S. Senanayke, J. Evans, A. Kubacka, **M. Estrella**, J. C. Hanson, A. Martínez-Arias, M. Fernández-García and J. A. Rodriguez. *High Activity of $Ce_{1-x}Ni_xO_{2-y}$ in Ethanol Steam Reforming: Tuning Catalytic Performance through Metal-Oxide Interactions.* Angewandte Chemie-International Edition, 2010. **49**(50): p. 9680-9684.
7. L. Barrio, A. Kubacka, G. Zhou, **M. Estrella**, A. Martinez-Arias, J. C. Hanson, M. Fernandez-Garcia and J. A. Rodriguez. *Unusual Physical and Chemical Properties of Ni in $Ce_{1-x}Ni_xO_{2-y}$ Oxides: Structural Characterization and Catalytic Activity for the Water-Gas Shift Reaction.* Journal of Physical Chemistry C, 2010. **114**(29): p. 12689-12697.
8. S. D. Senanayake, D. Stacchiola, **M. Estrella**, L. Barrio-Pliego, J. Evans, M. Pérez, J. Hrbek and J. A. Rodriguez. *Probing the reaction intermediates for the Water-Gas Shift over inverse CeO_x/Au(111) catalysts*. Journal of Catalysis, 2010. **271**(2): p. 392-400.

9. L. Barrio, **M. Estrella**, G. Zhou, W. Wen, J. C. Hanson, A. B. Hungria, A. Hornés, M. Fernández-García, A. Martínez-Arias and J. A. Rodriguez. *Unraveling the active site in copper-ceria systems for the Water Gas Shift reaction: in-situ characterization of an inverse powder CeO_{2-x}/CuO-Cu catalyst.* Journal of Physical Chemistry C, 2010. **114**(8): p. 3580-3587.
10. A. Hornés, A. B. Hungría, P. Bera, A. López Cámara, M. Fernández-García, A. Martínez-Arias, L. Barrio, **M. Estrella**, G. Zhou, J. A. Fonseca, J. C. Hanson and J. A. Rodriguez. *Inverse CeO_2/CuO catalyst as an alternative to classical direct configurations for preferential oxidation of CO in hydrogen-rich stream.* Journal of the American Chemical Society, 2010. **132**(1): p. 34-35.
11. **M. Estrella**, L. Barrio, G. Zhou, X. Wang, Qi Wang, W. Wen, J. C. Hanson, A. I. Frenkel and J. A. Rodriguez. *In Situ Characterization of $CuFe_2O_4$ and Cu/Fe_3O_4 Water-Gas Shift Catalysts.* Journal of Physical Chemistry C, 2009. **113**(32): p. 14411-14417.

Communications- (* indicates presenter)

March 2015	2015 JHU Biochemistry and Molecular Biology Retreat • Poster Presentation: "Characterization of the *Thermotoga maritima* Cmr Complex Demonstrates a Distinct Targeting Mechanism" **M. Estrella*** and S. Bailey	Towson, MD
March 2014	2014 JHU Biochemistry and Molecular Biology Retreat • Poster Presentation: "Putative roles for the Cmr1 and Cmr5 subunits of the *Thermotoga maritima*" **M. Estrella,** F. T. Kuo*, and S. Bailey	Towson, MD
June 2013	2013 CRISPR Meeting • Poster Presentation: "Characterization of the *Thermotoga maritima* Cmr Complex Demonstrates a Distinct Targeting Mechanism" **M. Estrella*** and S. Bailey	St. Andrews, UK
May 2013	2013 JHU Biochemistry and Molecular Biology Retreat • Oral Presentation: "RNA-guided RNA Cleavage by the *Thermotoga maritima* Cmr Complex Demonstrates a Distinct Targeting Mechanism" **M. Estrella*** and S. Bailey	Towson, MD
Apr 2012	2012 JHU Biochemistry and Molecular Biology Retreat • Oral Presentation: "Structural and functional characterization of the *Thermotoga maritima* Cmr complex" **M. Estrella*** and S. Bailey	Towson, MD
Sept 2011	20th JHU Institute for Biophysical Research Retreat • Poster Presentation: "CRISPR: The Prokaryotic Defense System" H. Chen*, **M. Estrella***, S. Mulepati and S. Bailey	Towson, MD
June 2009	21st North American Catalysis Society Meeting • Oral Presentation: "Structural and spectroscopic study of the WGS activity for the $CuFe_2O_4$ catalyst" **M. A. Estrella***, L. Barrio, G. Zhou, W. Wen, X. Q. Wang, J. C. Hanson and J. A. Rodríguez	San Francisco, CA

May 2009	Frontiers in Catalysis	Annapolis, MD

- Poster Presentation: "*In situ* measurement of metal oxide and metal rearrangement of $CuFe_2O_4$ spinel during catalyst activation for Water Gas Shift (WGS) reaction catalysis" J. C. Hanson*, L. Barrio, **M. Estrella** ,W. Wen, X. Q. Wang, G. Zhou and J. A. Rodriguez

Nov 2008	Northeast Reg. Meeting of NOBCChE at St. Francis College	Brooklyn, NY

- Oral Presentation: "In situ studies of CeO_{2-x}/CuO_{1-y} catalyst during Water-Gas Shift reaction using XRD and XANES" **M. A. Estrella***, L. Barrio, W. Wen, J. C. Hanson and J. A. Rodriguez

Nov 2008	Oak Ridge Nat. Lab., Science & Energy Research Challenge	Oak Ridge, TN

- Poster Presentation: "In situ studies of CeO_{2-x}/CuO_{1-y} catalyst during Water-Gas Shift reaction using XRD and XANES" **M. A. Estrella***, L. Barrio, W. Wen, J. C. Hanson and J. A. Rodriguez

Oct 2008	Stony Brook Univ., 9th Annual Chemistry Research Day	Stony Brook, NY

- Poster Presentation: "In-situ Time-Resolved X-ray Diffraction and X-ray Absorption Spectroscopy for the Characterization of Catalysts" J. C. Hanson*, L. Barrio, **M. Estrella***, W. Wen, X. Q. Wang and J. A. Rodriguez

Aug 2008	Brookhaven Nat. Lab., Chemistry Dept. Presentation	Upton, NY

- Oral Presentation: "In situ studies of CeO_{2-x}/CuO_{1-y} catalyst during Water-Gas Shift reaction using XRD and XANES" **M. A. Estrella***, L. Barrio, W. Wen, J. C. Hanson and J. A. Rodriguez

Apr 2007	NOBCChE National Conference	Orlando, FL

- Oral Presentation: "Recombinant Protein Purification by Self-Cleaving Aggregation Tags" **M. A. Estrella***, W. Y. Wu, C. Mee, F. Califano, R. Banki and D. W. Wood

Invited Talks

February 9, 2016	Johns Hopkins University, Biophysics Department	Baltimore, MD

- "CRISPR-Cas: Why is it an important discovery?" **M. Estrella*** and S. Bailey

Dec 6, 2012	St. Francis College, Chemistry Department	Brooklyn, NY

- "Structural and biochemical characterization of an RNA-guided prokaryotic immune system" **M. Estrella*** and S. Bailey

Activities and Service

March 2014 & 2015	Student organizer of the Biochemistry and Molecular Biology Retreat
Jan 2013 – Jan 2016	Member, American Academy of Arts and Sciences
June 2012 – 2015	Speaker for Meyerhoff Scholars Program JHSPH visit
Sept 2012 – Sept 2014	PhD student representative, BMB Department